1 of 2
AD A
103395

AGARD-CP-311

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE 92200 NEUILLY SUR SEINE FRANCE

AGARD CONFERENCE PROCEEDINGS No. 311

Aural Communication in Aviation

NORTH ATLANTIC TREATY ORGANIZATION

**DISTRIBUTION AND AVAILABILITY
ON BACK COVER**

81 8 27 038

NORTH ATLANTIC TREATY ORGANIZATION

ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT

(ORGANISATION DU TRAITE DE L'ATLANTIQUE NORD)

AGARD Conference Proceedings No.311

AURAL COMMUNICATION IN AVIATION

Edited by

Dr K.E.Money
Defence and Civil Institute of
Environmental Medicine
1133 Sheppard Avenue West
P.O.Box 2000
Downsview, Ontario M3M 3B9
Canada

THE MISSION OF AGARD

The mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

Exchanging of scientific and technical information;

Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;

Improving the co-operation among member nations in aerospace research and development;

Providing scientific and technical advice and assistance to the North Atlantic Military Committee in the field of aerospace research and development;

Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field;

Providing assistance to member nations for the purpose of increasing their scientific and technical potential;

Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

꘠

PREFACE

The importance of aural information in aerospace operations is secondary only to visual information yet, despite the dependance of military operations on reliable voice communication and the effective use of audio warnings, many of the systems currently employed have serious shortcomings and do not reflect the considerable research effort that has been expended in this area.

The Aerospace Medical Panel (AMP) of AGARD, in recognising this fact, considered it to be timely to discuss this topic, so that there could be a wider understanding of the problems of speech recognition in aircraft noise, the causal mechanisms of subsequent aviator hearing loss and of the latest developments in research to combat the shortcomings in voice communications systems.

The papers which follow were presented by medical and scientific experts at the AMP Specialists' Meeting at Soesterberg airbase, Netherlands, from 30 March 1981 to 2 April 1981.

J.M.MULLANEY

AEROSPACE MEDICAL PANEL

Panel Chairman: Médecin Colonel J.Bande, BE

Panel Deputy Chairman: Air Commodore P.Howard, UK

Panel Executive: Squadron Leader J.M.Mullaney RAF, UK

MEETING ORGANISATION

Session Organiser: Dr K.E.Money, CA

Host Nation Coordinator: Mr E.H.J.Bleeker, NL

SESSION ORGANISER

Dr K.E.Money
DCIEM
1133 Sheppard Avenue West
P.O.Box 2000
Downsview, Ontario M3M 3B9
Canada

SUB-SESSION CHAIRMEN

Dr K.E.Money   Mr R.Camp
(see above)   USAARL
                     P.O.Box 541
                     Ft.Rucker
                     AL 36360
                     USA

                     Col. G.C.Mohr
                     Commander
                     AFAMRL
                     Wright-Patterson AFB
                     OH 45433
                     USA

CONTENTS

† This paper was not presented at the meeting.

TECHNICAL EVALUATION REPORT
AURAL COMMUNICATIONS IN AVIATION

by K.E. Money

Defence and Civil Institute of Environmental Medicine
Downsview, Ontario, Canada

The presentations and discussions concerning aural communications in aviation reveal two major problems (opportunities). One of these is that the communications systems are basically inadequate and notoriously variable so that voice communication, except for routine transmissions, is frequently difficult and slow. Often it is on those occasions when it is especially important that auditory information be received with ease that communication is difficult, but even routinely the poor quality of the equipment adds unacceptably to the workload of the aircrew.

The second major problem/opportunity regarding aural communication in aviation is that aircrew frequently suffer preventable noise-induced hearing losses. This second problem is much intertwined with the first problem, of course, since aircrew with greater hearing losses will have greater difficulty in perceiving auditory communications with inadequate equipment and the equipment itself generates much of the noise that causes the hearing losses. Noises from engines and weapons systems also contribute to hearing losses, but regardless of how a pilot has acquired his hearing loss (from recreational noise exposure, from small arms training, or from his aviation environment) a pilot with a hearing loss will probably use a higher "volume" setting to his ear phones while flying in order to hear, and the resulting greater intensity of clicks, snaps, static, noise, and voices will then speed up the progression of his hearing loss.

Most armed forces have criteria whereby trained (usually older) aircrew are considered unfit for flying if hearing losses grow too large. These criteria are not unreasonably conservative, and it is fair to say that most people who exceed these criteria have acquired a social handicap. However, when these criteria are exceeded the individual is almost always given a waiver allowing him to continue to fly, because it can be shown that he "can still do his job safely". It is perhaps a fair guess that in most cases he can still do his job because of long experience and a very high volume setting to his ear phones. Clearly, more effort should be given to the determination of whether the job is still exacerbating the hearing loss.

In addition to the provision of high quality communications systems and the prevention of noise-induced hearing losses, other opportunities in this area include voice communications security and jamming, and development of voice-activated controls. The presentations describe in some detail how to benefit from these opportunities.

The presentations also show that in many respects our present state of knowledge in this area is obviously far ahead of our ability to implement it. We are using dynosaur-age communications systems in our space-age aircraft, with costly results, although superior equipment is well within the state-of-the-art. It is clear even that this has been the case for many years. The difficulty is, perhaps, that the provision of a high quality communications system, and the provision of protection from noise, have not been the clear responsibility of any one person or any one group. The flight surgeons are not the engineers, the engineers are not life scientists with knowledge of auditory physiology and audiology, the life scientists find it difficult to get the ear (!) of the operational managers, and the aircrew have admirable tendencies to refrain from complaining and to see difficulties as obstacles to be overcome by individual efforts. Nevertheless, the cooperation of all these groups will be recruited easily when it is understood how inexpensive it can be to acquire high quality communications systems, with effective protection from noise, and with impressive benefits to operational effectiveness, to flight safety, and to the health of the aircrew. In general, the welfare of the aircrew is a responsibility of all of these groups, but it is primarily the responsibility of the managers (operational commanders) and it is from them that the effective initiative should come after they have acquired the information to be found here.

For the systems that can be built now, and also for future systems, knowledge and research in the following areas are implicated:

a. **Physical Characteristics of Speech**

   Spectrum and dynamic range and actual content of vital information of the speech signal. Requirements to reproduce speech signals so that the intelligibility of the message, voice identification and other subtleties that are used in the fast and efficient communication process are not degraded.

b. **Ideal Systems Design for Transmission and Reception of Speech**

   Characteristics of microphones, pre-amplifiers, transmitters, receivers, intercommunication systems, earphones and earcups. System performance and effect of distortion (e.g. frequency, harmonic, peak-clipping) on speech intelligibility in normal and high noise environments.

c. **Speech Intelligibility Tests**

   Validity of speech tests that are employed in system evaluation. Problems of developing tests that are sensitive to the effects of distortion and that will measure other subtleties of speech. Examination of how present inadequate systems were tested.

d. Cost of Inefficient Systems

Damage to hearing from high acoustic energy of communication systems. Significance of failure or inadequacy of speech communication or audio warning systems in military operations -- cost in training and reduction of operational effectiveness.

e. Evaluation of Standard Military Electronic Speech Communication Systems

Comparison of present systems with the ideal application of state-of-the-art techniques. Procedures for the improvement of standards.

f. Audio Warnings and Controls

Voice versus tone warnings. Design of highly descriminative audio warnings. Optimum number of warnings to be employed. Relative merits of realistic versus synthesized speech in voice warnings, word content and word order. Relation of audio warnings to cockpit voice levels. Use of audio warnings in conjunction with visual displays. Voice activated controls. Voice communications jamming.

g. Hearing Standards

Relation to occupational task.

h. Hearing Conservation

Aircrew and ground personnel.

We shall have discharged our responsibilities when the equipment now readily available is installed in our aircraft, and when knowledge and research in these areas are being applied and pursued for the benefits and savings that they can produce.

VOICE COMMUNICATION RESEARCH AND EVALUATION SYSTEM

Richard L. McKinley
Air Force Aerospace Medical Research Laboratory
Wright-Patterson Air Force Base, Ohio 45433

SUMMARY

Aircraft voice communications may be degraded by a variety of sources such as electrical and/or acoustical noise, radio interference, jamming and various other forms of distraction. The Voice Communication Research and Evaluation System, located in the Biodynamics and Bioengineering Division of the Aerospace Medical Research Laboratory, has been developed for the comprehensive analysis and enhancement of operational voice communication. The basic system is comprised of a multi-station voice communication network consisting of the USAF standard aircraft intercommunication system, a standard A-19 diluter-demand oxygen regulation system and an on line computer data collection and data analysis system that displays results in real time. The system is housed in a large reverberation chamber containing a programmable sound source capable of reproducing the spectrum and level of any AF operational noise environment. Standardized voice communication effectiveness test materials are used to assess the performance of any aspect of the total voice communication link, however, emphasis is usually placed upon the performance of the aircrew members. This paper will describe the salient features of this unique system and provide examples of its application to voice communication problems.

INTRODUCTION

Air and ground crew voice communications may be degraded by a variety of system and environmental factors that include electrical or acoustical noise or both, radio interference, jamming, communication signal processing and various other factors that prohibit effective communication. Vigorous research activities must be maintained to identify and quantify elements that cause such deterioration and to develop principles, techniques and guidelines that will minimize adverse effects and optimize voice communications. Analytical studies of communication system performance, environmental influences and the man-in-the-loop element must be carried out under carefully controlled conditions that simulate to the greatest extent possible, the practical operational situations of concern. Such efforts are possible in controlled laboratory environments where special instrumentation can be used to create the essential elements of human factors and communication system networks being investigated.

A Voice Communication Research and Evaluation System (VOCRES), located in the Biodynamics and Bioengineering Division of the Air Force Aerospace Medical Research Laboratory has been developed to provide the capability for comprehensive research, test and evaluation activities in the voice communications effectiveness arena. VOCRES, the subject of this report, has been designed to replicate system and environmental variables believed to have significant influence on operational communication. Using VOCRES, various elements of voice communications can be analyzed either individually or in component clusters. Using this method of analysis, problem areas can be identified, attacked and the overall operation enhanced. The effectiveness of various interfering signals may also be evaluated by their insertion into the communication system. The operational procedures and materials used in the laboratory are well standardized and provide data with a high degree of reliability.

This report describes the VOCRES system instrumentation in some detail as well as the psychoacoustical procedures used in the overall operation of the voice communication research program. The key element of the overall program is VOCRES. Other component systems are essential to the realistic replication of communication situations for expanding the technology base as well as performing discrete measurements required for the treatment of specific problems.

APPROACH

The general approach employed in this program involves the participation of volunteers who communicate as talkers and listeners under controlled conditions that replicate the specific communication environments being evaluated. Subjects are stationed at custom-designed consoles and communicate with standardized or special purpose (speech) vocabulary materials while various system and environmental characteristics or equipment are varied and the resulting communication effectiveness is quantified. Elements commonly varied are microphones, earphones, ambient noise level at the crew station, helmets and oxygen masks, aircraft radios, jamming signal type and modulation, jammer to signal power ratios, and receiver input power. Data derived from these efforts may be used to establish baseline communication system performance profiles, for comparative testing of specific communication system components, such as radios, intercoms, microphones, earphones, and voice processors. The data are also used to quantify the performance of a specific component in a specific environment. Subjective comments from active aircrew personnel who have experienced the VOCRES reveal that the validity of the system and of the approach is quite good.

INSTRUMENTATION

VOCRES: General System

The VOCRES system is an aggregate of four different subsystems integrated into a voice communication network that includes ten individual communication stations and one control station. The individual communication consoles are located in a large reverberation chamber and the master console is located in a control room adjacent to the chamber. The general physical assemblage of the individual subsystems and the integrated system is displayed in Figure 1.

LEGEND
A POWER SUPPLY
B SPEAKER BANKS (8)
C 7 KW LOW FREQ. AMP.
D 7 KW HIGH FREQ. AMP.
E 600 W LOW FREQ. AMP.
F 600 W HIGH FREQ. AMP.
G SOUND SYSTEM CONTROL CONSOLE
H AIRCRAFT RADIO (AN-ARC-164)
I HP 9845 COMPUTER
J X-Y PLOTTER
K DISK DRIVE
L BREATHING AIR SUPPLY
—— AIR LINES
––– COMPUTER LINES
--- INTERCOM LINES
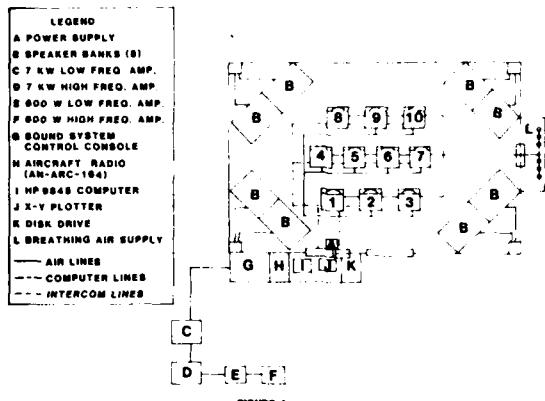
FIGURE 1

The subsystems include (1) an AIC-25 aircraft intercommunication system (11 stations), for use with Air Force standard communications headgear, (2) an air respiration system with A-19 diluter-demand regulators for use with standard oxygen masks, (3) a high intensity sound source for duplicating operational acoustical environments occupied by crew members and (4) a central processing unit that controls all stations and conducts the individual testing sessions and conditions, i.e., presents materials, monitors participant activity, records, stores and analyzes responses, and provides analyzed data in tabular or graphic form or both. The overall system is adaptable to the incorporation of various aircraft radios, communication jammers, and the like, that are not integral components of VOCRES.

Each of the ten communication consoles or stations is equipped with an AIC-25 intercommunication terminal, an A-19 respiration terminal, a display/subject response unit, a keyboard for communication performance task response from the participants and a large volume unit (VU) meter that indicates voice level of communications generated at that station (see Figure 2). The system can be operated with any number of one to ten volunteers. The psychophysical paradigm used most often is a "round robin" procedure where each subject, in turn, performs as a talker while the remaining subjects respond as listeners.
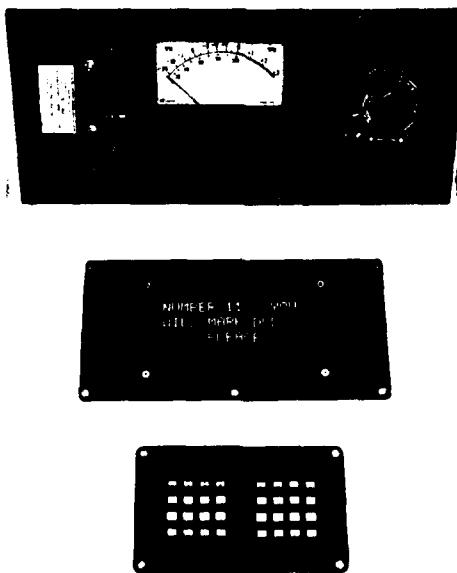
FIGURE 2

Communication Materials

Communication materials consist of the standardized Modified Rhyme Test* for most activities with VOCRES. Other test materials such as the Diagnostic Rhyme Test are used from time to time for special purpose applications.

Communication Link Capabilities

The communications assemblage diagram (Fig. 3) demonstrates the high flexibility of VOCRES that allows a variety of different communication links to be examined either individually or in combination with one another. The range of communication links can be varied from a simple face-to-face communications situation (i.e., direct talker to listener) to a complex configuration using encoders, encrypters, and the like by varying appropriate subunit controls. Any of the alternate pathways shown in Figure 3 can be used to complete the talker to listener link. The direct talker to listener path theoretically provides a data baseline free from environmental and component effects. The influences of the various elements of the communication system operation relative to the baseline can be quantified, analyzed and evaluated by measuring performance while varying single components and clusters of components of the VOCRES.
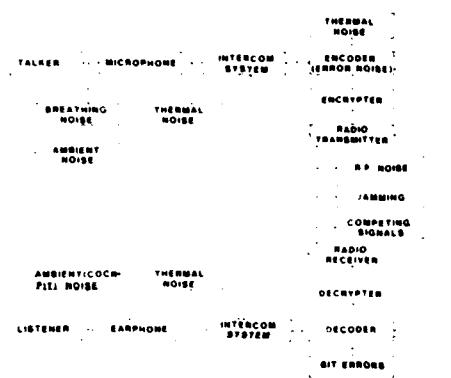
FIGURE 3
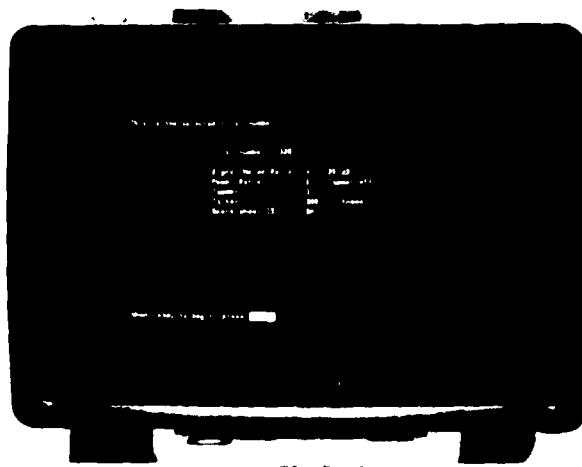
FIGURE 4

Central Processor-Display-Response System

The control console of the system includes a typewriter type keyboard and a cathode ray tube (CRT) display. Through this console the test administrator enters the required experimental information. The central processing unit then displays the required instrument on the CRT (Fig. 4). After all experimental instrument settings are completed and stablized, the administrator tells the central processing unit to administer the selected test and collect data from each of ten individual communications desks. The system is capable of making any one of the 10 stations the talker position and also can facilitate multiple talkers. For example, during a test one subject will be designated a talker for a list of 50 words and the other nine subjects will be designated as listeners. On the CRT the system displays each of the listeners' responses to each item spoken by the talker. The CRT display also indicates whether or not the response is correct (Fig. 5). The central processor-display-response system is diagrammed in Figure 6. The central processor is the Hewlett-Packard 9845T System. This system has dual 16-bit pro-

FIGURE 5

*Standardized lists of 50 monosyllabic words; each list developed to be essentially equivalent in intelligibility to the other lists.

cessors, wherein one handles internal functions, while the second handles I/O functions. Also included in the basic system are a CRT with graphics, a thermal line printer (8-1/2" wide) with graphics capability and two cartridge tape drives, each with 217K byte* capacity. A 20 Mega byte disk drive with two platters, one fixed and one removable, adds additional data storage capability. Several interfaces are also included in the system. An RS-232 interface is used for sending and receiving data from the individual communication stations, while an IEEE 488-1975 General Purpose Interface Bus is used for control and data collection from various electronic instruments. These include a digital spectrum analyzer, a frequency synthesizer, a digital voltmeter, an RF power meter, and a 4-color flat bed X-Y plotter. A second RS-232 interface receives data from a digital oscilloscope or an audio tape deck.
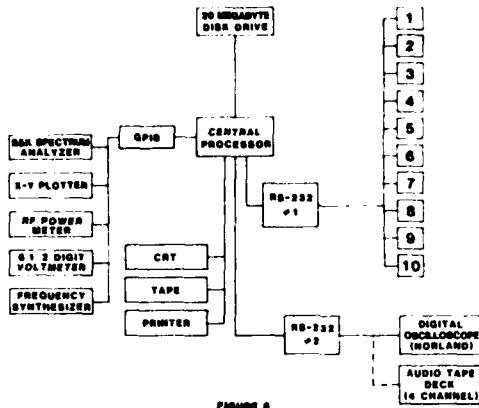
FIGURE 6

Each individual communication desk has its own RS-232 compatible interface shown in Figure 6 which decodes commands by the central processor for the display system and also returns the subjects' responses to the central processor for storage and analysis. Each desk station interface has two addresses to which it will respond. One address is common to all desks, therefore by using one address and message, all desk displays can be activated or loaded simultaneously. The second address is specific to only one desk and by using this address, ten different messages can be loaded into each of ten different displays. The interface for each desk operates at 9600 bits per second allowing seemingly simultaneous operation at each of the ten stations.
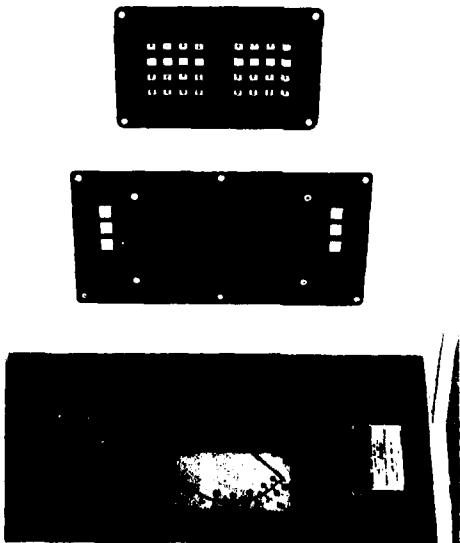
FIGURE 7

*Note: 1 byte = 8 bits

Figure 7 shows one of the 64 character alphanumeric gas discharge, type displays. Each character is 5.73 mm (.023 in) x 8.27 mm (0.33 in) and is generated by a 5x7 dot matrix with a separate underline capability. The display is very bright having a level of 30 ft-L. The contrast of the neon-orange characters is enhanced by the use of a circularly polarized filter.

The subjects can respond by using one of two different response systems. The first system consists of six pushbuttons, three on either side of the displays each with a red LED mounted in the bezel. Pressing one button causes the adjacent LED to light indicating a response has been made. Pushing a second button will allow the volunteer to change his decision, illuminating the second light instead of the first. The second response system consists of two 4x4 calculator type keypads. Only one of the 32 buttons can be chosen at one time. Operation is similar to the six LED pushbuttons except that pressing one of the keys causes from one to five of the six LEDs to light forming a specific pattern for that key. These LEDs provide feedback to the subject indicating the chosen response.

Data Treatment

Computer software was developed to standardize test procedures and to facilitate the administration of the Modified Rhyme Test or any other standardized intelligibility test over a large number of individual trials. The software also includes the experimental design. Each test parameter is displayed on the CRT before the trial and appropriate equipment settings are made by both the test administrator and central processing unit. The individual units of the Modified Rhyme Test or any other test materials are stored on the system's 20 Mega Byte hard disk. Following each trial, data for each subject, all test parameters and the time of the trial are stored on the system's disk. Fail-safe backup is accomplished by printing the same data on the system's thermal line printer. The data may be analysed at any time, using a variety of standard statistical measures and plotting techniques. This method of data storage and analysis can give preliminary results in real time.

High Intensity Sound System

The high intensity sound system is shown in Figures 8 and 9. The system is capable of operation in one of two power modes, a high power mode where 14,000 watts are available and a low power mode where 1,200 watts are available. The power amplifiers drive eight banks of loudspeakers containing a total of 96 Altec 15" low-frequency speakers, eight Altec horn loaded compression drivers, and 384 Stromberg Carlson high frequency speakers. The noise generator and the spectrum shaper allow almost any desired noise environment (spectrum) within the human audio-frequency range to be generated inside the test chamber. This permits the accurate reproduction of ambient and environmental noise conditions of specific operational situations within the laboratory, which is a vital aspect of the validity of the communication testing.
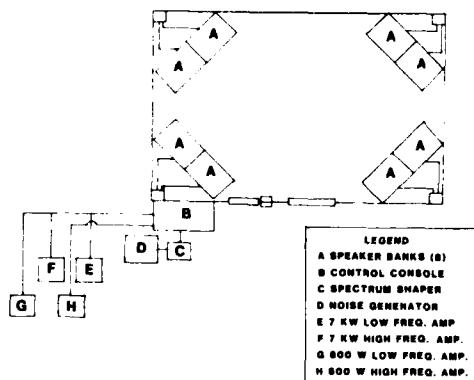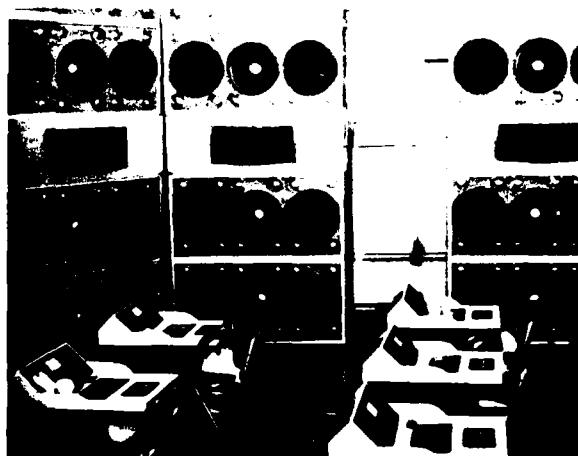
FIGURE 8   FIGURE 9

The room in which the loudspeaker banks are located is a specially designed and constructed acoustic reverberation chamber. The room is designed for maximum reverberation time and approximately 8000 ft^3 in volume. The irregular wall surfaces are designed to disrupt the formation of standing waves and maximize the uniformity of the level of a noise distributed throughout the room.

AIC-25 Intercommunication System

The aircraft intercommunication system shown in Figure 10 is a standard AIC-25 intercommunication system. The test administrator and each desk has an individual AIC-25 aircraft intercommunication unit. A switching circuit located on the control console allows the talker's intercom to be disconnected from the rest of the system and taken directly to the audio input of any transmitter. The audio output of the receiver is then routed to the other nine listeners. The terminal equipment available for the intercom system includes standard H-157A headsets, H-133 headsets, MBU-5/P oxygen masks, and HGU-26/P flight helmets. A sample of each of these is pictured in Figure 11.
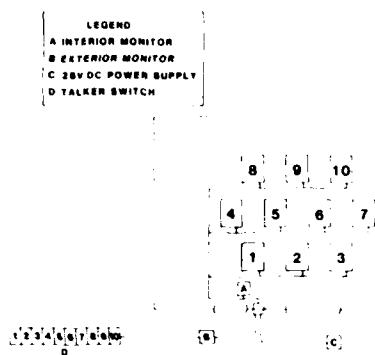
FIGURE 10   FIGURE 11

Air Respiration System

The air breathing system depicted in Figure 12 uses the standard Air Force A-19 diluter demand regulator as the primary item in the system. Each station has its own A-19 regulator which is supplied through feeder lines by a semiautomatic regulator manifold. The manifold connects six standard size breathing air bottles to the system through two regulators. Each regulator controls three bottles. When the supply of the first three bottles is exhausted the system automatically switches to the second set of three bottles. The normal operating pressure in the system is 150 psig.

FIGURE 12

Each of the above systems is integrated into each of ten individual subject stations. The final product is shown in Figures 13 and 14. The desk was designed for minimum size to minimize acoustical reflections from the surface and yet be functional. Each station is independent.
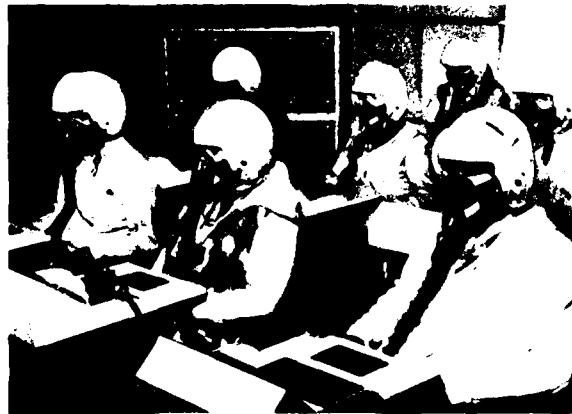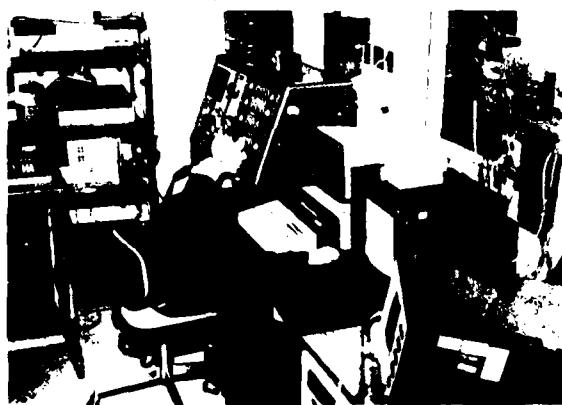
FIGURE 13   FIGURE 14

In the past, interim versions of the VOCRES system were used to evaluate communication properties of lightweight helmets, chemical defense ensembles, new oxygen masks, and innovative radio systems. Current studies involve the investigation of effects of jamming on communication in a quantitative manner relative to the J/S, S/N, radio type and jammer type, evaluation of new chemcial defense ensembles, and development of new communication microphones. Future studies will include modeling of human response to jamming and enhancement of terminal communication equipment.

CONCLUSION

This paper has described the VOCRES system, its capabilities and uses. In conclusion, VOCRES is a semiautomatic laboratory voice communication test system that uses human subjects in a realistic communication environment to conduct research, test and evaluation of Air Force communications systems and their effectiveness.

REFERENCES

House, A. S., Williams, C., Hecker, M. H. L., and Kryter, K. D. Psychoacoustic Speech Tests: A Modified Rhyme Test, ESD-TDR-63-403, Electronic Systems Division, L. G. Hanscom Air Force Base, MA, 1962.

Voiers, W. D. Performance Evaluation of Speech Processing Devices. III. Diagnostic Evaluation of Speech Intelligibility. AFCRL-67-0101, Air Force Cambridge Research Laboratories, L. Y. Hanscom AFB, MA, 1967.

VOICE COMMUNICATIONS JAMMING RESEARCH

Thomas J. Moore, Ph.D.
Air Force Aerospace Medical Research Laboratory
Wright-Patterson Air Force Base, Ohio 45433

SUMMARY

This paper describes some of the current research on voice communication jamming being conducted within the Biological Acoustics Branch of the US Air Force Aerospace Medical Research Laboratory. The primary study described evaluated the effect of various types of jammers, J/S power ratios and background noise levels on voice communication materials processed by a standard Air Force voice communication system. The measurement instruments employed were the Modified Rhyme Test (MRT) and a non-standard voice communication performance task. Good agreement was found between the two measurement instruments in ranking the relative effectiveness of the jammers evaluated.

Summary data is also presented from a recent exploratory study in our lab which investigated the effect of training on naive subjects' ability to listen to voice communications under conditions of noise and jamming. The results of this exploratory study indicated that training improved the performance of the listeners under all conditions tested. The results also indicated that a comprehensive study is needed to quantify such factors as (1) jammer specific performance, (2) fatigue effects, (3) individual listener differences.

INTRODUCTION

Sustained effective voice communication is a vital requirement in modern tactical air warfare (1). Because of this, increased emphasis is being placed on the use of electronic jamming techniques against voice communications and on the development of special purpose radio systems with increased jam resistance capabilities. While much is known concerning the relative effectiveness of various types of jamming against specific radio systems given particular parameters, e.g. jammer to signal (J/S) ratio, antenna configuration, distance and atmospheric effects, these results are based to a great extent on equipment vs equipment tests. On the other hand, experience at field exercises (EW/CAS, Red Flag) has demonstrated that there exists a large grey area where a specific result predicted from equipment vs equipment tests is not obtained when a human operator becomes part of the loop. The Biological Acoustics Branch of AFAMRL has started to address the problem of what influence do various human and selected environmental factors have on the overall effectiveness of particular jamming situations and this paper will describe some of the current research within our laboratory.

The initial attempt to address this problem consisted of a two phased study. The first phase measured the comparative intelligibility of standardized word test materials processed through typical Air Force aircraft voice communication systems in the presence of simulated operational noise and different types and levels of jamming. The second phase measured the same system parameters and used a non-standard communication performance task as the measurement instrument instead of standardized test materials. This task was conceived to provide a more "realistic" test of jamming effectiveness than the standard word test. The questions addressed by this study included:

*Over a range of jammer to signal power ratios does the presence of simulated operational noise differentially affect the intelligibility of words processed over a standard Air Force aircraft voice communication system?

*What is the comparative effectiveness of different types of jammers?

*Are the results yielded by the more "realistic" communication task comparable to those from the standardized test materials?

METHOD

Approach

The comparative intelligibility of standardized test materials and of a voice communication task processed through a representative Air Force voice communication system was measured in the presence of jamming and simulated operational noise. Volunteer listeners wearing standard inflight helmets and oxygen masks responded to the communication signals under the specified experimental conditions. Decrements in comparative intelligibility were attributed to type and level of jamming employed and level of simulated operational noise.

Subjects

Ten volunteer subjects, five male and five female, were employed in the present studies. All were recruited from the general civilian population. They were paid at an hourly rate for their participation, with a cash bonus awarded when the subject completed all sessions. The hearing levels of all subjects were no greater than 15 dB at any standard audiometric test frequency from 500 to 6000 Hz.

Facilities

These studies employed the Voice Communication Research and Evaluation System (VOCRES) of the Aerospace Medical Research Laboratory (2), which will be described more fully in a paper to be presented later during this meeting. This system has the capability to realistically model the major acoustic factors experienced by crew members that may adversely affect voice communications.

The overall system includes a master control station and ten individual aircraft communication stations. Each station contains the Air Force standard intercommunication system (AIC-25) and respiration system (A-19 Oxygen Regulators) as well as appropriate terminal equipment, i.e., headsets, helmets, oxygen mask-microphones, etc. Both intercommunication and respiration terminals and operating controls are easily accessible to the individual positioned at the station. Each station is also integrated into a Computer-Display-Response system in which the central processor is a Hewlett-Packard 9845T System. An interface at each station decodes commands by the central processor for the station's display and also returns the subject's response to the central processor for storage and analysis.

For these studies all volunteers wore the standard Air Force flight helmet, HGU-26/P with the H-154A earcup assembly. Helmets were individually fitted to each subject by personnel of the AFAMRL Human Engineering Division. Either the MBU-5/P or MBU-12/P AF standard oxygen mask with the M-101 noise cancel-ling microphone was worn by each volunteer. Previous evaluations have determined that the communication performance of these two masks is equivalent. Compressed air was respired through A-19 Diluter Demand Pressure Breathing Regulators set at normal operation for all subjects during the talking and listening phases of the studies.

The acoustic environment simulation facility consists of a large reverberation chamber (approximately 8000 ft^3) that houses a powerful electrodynamic sound system. The electrodynamic system contains dual amplifiers that may be used singly or in combination. One system (low power) consists of two 600 - watt amplifiers and the other (high power) consists of two 7000 - watt amplifiers. The amplifiers drive 32 loudspeaker enclosures, each containing three 15-inch loudspeakers and twelve 3-inch high frequency "tweeters," eight additional enclosures contained high frequency compression drivers with horns. The loudspeaker enclosures are portable and may be rearranged for various purposes. In the configuration used for these studies, the low power system generates a maximum overall Sound Pressure Level (SPL) of 122 dB, while the high power system generates a maximum overall SPL of 128 dB re 20 μPa (with a pink noise input).

The low power system was used to simulate the cockpit noise environment of tactical fighter aircraft. A pink noise source was shaped by a 1/3 octave band spectrum shaper (or filter bank) so that the spectrum measured in the test space was representative of that produced by a typical jet fighter aircraft.

The transmitter and receiver were ARC-164 radios. The ARC-164 is the current operational Air Force aircraft radio. The communication signals were processed by the radios and then presented to the listeners through the standard Air Force aircraft intercommunication system and terminal equipment.

The types of jamming signals and the jammer to signal ratios that are reported here were specified in cooperation with the Air Force Electronic Warfare Center (AFEWC) and are as follows:

(1) A FM tone
(2) A FM drifting tone
(3) A white noise AM modulated on an RF carrier
(4) Pulsed AM noise
(5) Dual FM swept tones

These jammers represent a subset of the total number of jammers investigated to date and were chosen to represent the range of effectiveness found. For some of these jammers simulated cockpit noise levels of 95 and 115 dB were investigated, in addition to the 105 dB and ambient chamber noise levels reported in this paper. The J/S ratios were -6 dB, -3 dB, 0 dB, +3 dB and +6 dB. All RF connections between the transmitter, receiver and jammer were made by means of standard 50 ohm coaxial cable, thus allowing precision step attenuators to be used to vary the J/S power ratio.

Measurement Instruments

The major difference between the two phases of this study was the measurement instrument employed. Phase I employed a standardized measure of intelligibility, the Modified Rhyme Test (MRT) as developed by House, et al (3) for assessing communication effectiveness. The MRT was selected for use over other test materials because of evidence that it is the test of choice for evaluating the performance of military speech communication systems in the presence of environmental noise (4). The materials consist of lists of 50 one-syllable words that are equivalent (lists) in intelligibility. These test words are presented embedded within a carrier phrase that is the same for each item. The MRT is easy to administer, score and evaluate and it does not require extensive training of listeners. MRT word intelligibility has been sufficiently standardized to allow the relative intelligibility of such materials as closed message sets and sentences to be estimated on the basis of the corresponding measured MRT scores.

During Phase I, the display at the talker's station provided the MRT text which the talker read. The listener's stations provided a choice of six MRT key words from which the listener selected the one he/she heard or believed was correct, by pressing the button beside it. The listener responses were auto-matically recorded, statistically analyzed and displayed by the central control unit. To compensate for correct answers obtained by guessing, a correction factor was applied to the scores during data analysis. This correction for guessing (# incorrect/# possible choices - 1) was subtracted from the total # correct for each subject.

Phase II used as the response measure a communication performance task that was designed to be more "realistic" in nature than the MRT. This task is an adaptation of one employed successfully by Ascher,

et al (5) to evaluate the ability to follow one out of three simultaneous voice messages in a simulated Naval Combat Information Center. In the present study the nine listeners were assigned to three groups, each containing three persons. Each of the groups was assigned a different call sign such as "Ringo," "Laker" or "Baron." The display at the talker's station provided the text that the talker read. This text provided one of the three call signs followed by instructions that designated a location in a 4x8 matrix, for example "Ready Ringo go to Red 4." The listener responded only to messages containing his/her call sign by pressing the appropriate button in the matrix at his/her station. Listener's whose call signs were not heard made no response. The task is scored in terms of the number of correct responses made. Errors made in any of the three elements of the message (call sign, row color, column number) are kept track of individually, an incorrect response in any one or all three of the key elements can be scored as one incorrectly received message because in an actual operational situation the desired action would not have been performed by the listener. When the data was analyzed we discovered that rarely, if ever, did a subject make a mistake with respect to the call sign. That is, practically all errors made were on the row or column designation. Therefore the results to be shown later are based upon only those 17 of the total 51 messages that contained an individual subject's correct call sign. Table 1 gives the key words used in the messages. Photographs of the displays for both the MRT and the communication performance task will be shown in McKinley's paper later in this meeting.

CALL SIGNS	ROWS	COLUMN
RINGO	RED	ONE
LAKER	BLUE	TWO
BARON	WHITE	THREE
	GREY	FOUR
		FIVE
		SIX
		SEVEN
		EIGHT

TABLE 1

Experimental Procedure

The present studies were designed so that each subject served as his/her own control, i.e., each subject participated in all experimental conditions. Order of presentation of experimental conditions was randomized, with the exception that only one type of jammer was evaluated during any one experimental session. Five of the ten subjects, three male and two female, were selected to serve as talkers in Phase I, while three of those five, two male and one female were used as talkers in Phase II. These individuals served as talkers and listeners in a "round-robin" fashion. The talker on any one list served as a listener on previous and subsequent lists. The parameters investigated were type of jammer, jammer-to-signal (J/S) power ratios and presence or absence of simulated cockpit noise. Subjects participated for 4 hours per day in experimental sessions of about 40 minutes followed by 15 minute rest periods. All ten subjects were run simultaneously within the VOCRES facility.

RESULTS

Phase I - Figure 1 displays communication intelligibility measured with the MRT for the five jamming signals at each of the jamming power ratios in the ambient noise condition of 79 dB, (the simulated cockpit noise exposure was not used for this first set of baseline data). The standard Air Force helmets and headphones worn by the subjects in all conditions provide an average sound protection of 20 dB. Similar measurements made in the simulated cockpit noise of tactical aircraft at a level of 105 dB are displayed as average values in Figure 2.
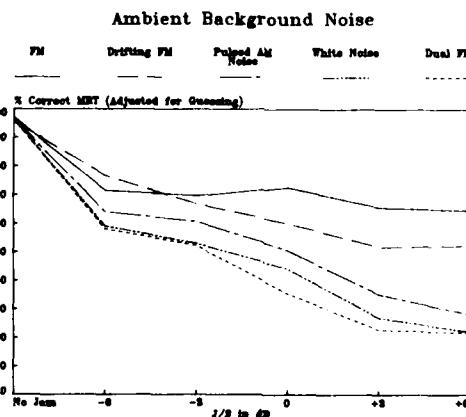
FIGURE 1

The results summarized in Figures 1 and 2 allow the following generalizations:

(1) In the absence of the simulated cockpit noise (Fig 1), four of the five jammers investigated result in decreased intelligibility as J/S ratio increases. The fifth jammer (FM tone) shows an initial effect of the presence of jamming with a decrease in intelligibility from 97% correct with no jamming to 71% correct with a -6 dB J/S ratio. Subsequent increases in J/S result in little further decrement in intelligibility for this jammer with a +6 dB J/S providing only an additional 7% decline in intelligibility.

When 105 dB of simulated cockpit noise is present, (Fig 2) there is a decrement of approximately 12% under the no jamming condition attributable to the presence of the noise. The presence of the background noise would appear to have the greatest effect upon the FM tone. The function for this jammer now shows a slight increase in effectiveness as J/S increases, with intelligibility at +6 dB J/S being 20% less than at -6 dB J/S.

105dB Background Noise
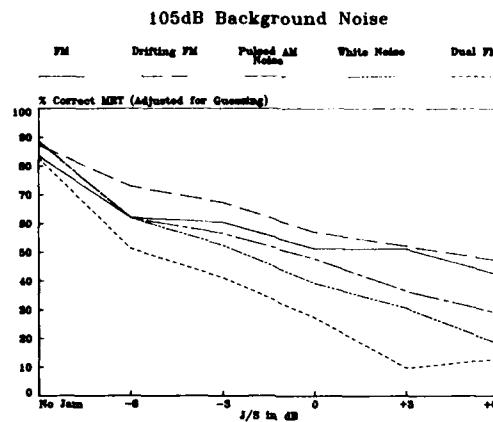
FIGURE 2

(2) Overall the Dual FM sweep tones appear to be the most effective jammer, with the AM modulated white noise the next most effective. Conversely, the FM tone and the FM drifting tone are the least effective.

Phase II - The results of the communication intelligibility measurements using the performance task described above as the response measure are summarized in Figures 3 and 4. Figure 3 shows data collected in the absence of simulated cockpit noise, while Figure 4 shows data collected when 105 dB of simulated cockpit noise was present. Examination of these figures indicates that, as was the case when the MRT was used as the response measure, the FM tone was the least effective and the Dual FM sweeping tones was the most effective of the five jammers shown.

Ambient Background Noise
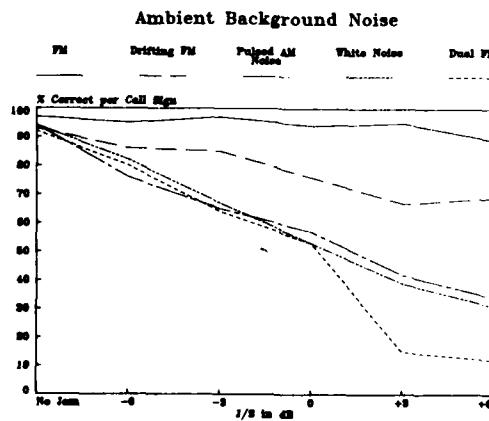
FIGURE 3

The similarity between results depicted in Figures 3 and 4 with those for the MRT (Figures 1 and 2) is striking. The trends and relative magnitude of effects is very similar for the two different response measures. The primary difference between the two response measures was that the presence of a jammer at the lowest power level (-6 dB J/S) did not cause as great a decrement in intelligibility relative to the no jamming condition when the response measure employed was the non-standard communication performance task as when it was the MRT. This is particularly apparent when the functions for the FM tone are considered. Also the relatively greater effectiveness of the Dual FM swept tones at +3 and +6 dB J/S is more apparent when the performance task is used as the response measure. Table 2 lists the correlation coefficients for a linear correlation between the scores on the MRT and the Performance Measure for each of the jammers investigated, both in the absence and the presence of the 105 dB simulated cockpit noise.

105dB Background Noise
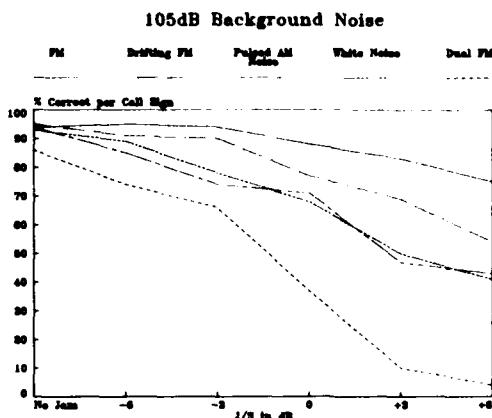
FIGURE 4

Exploratory Study on Training Effects in Jammed Voice Communications - The basic objective of this explora-
tory study was to investigate the performance of naive listeners given successive presentations of jammed
voice communications to determine whether there is a significant improvement in performance over time. The
laboratory facilities used were the same as described above.

LINEAR CORRELATION BETWEEN MRT AND
'COMMUNICATIONS PERFORMANCE' TASK

JAMMER	BACKGROUND NOISE LEVEL	
	79 dB (AMBIENT)	105 dB
FM Drifting Tone	0.94	0.90
White Noise	0.95	0.94
Dual FM	0.91	0.93
Pulsed AM Noise	0.98	0.94
FM	0.45	0.78

TABLE 2

The listeners responded to a total of 182 trials in this study with one MRT (50 items) presented
during each trial. The initial 41 trials were transmitted and received over ARC-164 radios, using the
AIC/25 intercom and standard helmets and earphones. On trial 42 a simulated tactical aircraft cockpit
noise environment at a level of 105 dB was added. On trial 99 an acoustic jamming signal at a +3 dB J/S
was added. The jammer used was the AM modulated white noise employed in the jammer evaluation study
described above. At the end of trial 182 the panel was dismissed. After a break of two weeks, two of the
six listeners and the two experienced talkers returned for further testing under the noise plus jammer
condition. Summary results of this exploratory study are shown in Table 3. A more detailed description
will be published in a report presently under preparation (6).

SUMMARY TABLE OF RESULTS OF TASK
"LEARNING" VS EXPERIMENTAL CONDITION

TEST CONDITION	PERFORMANCE - PERCENT CORRECT RESPONSE		PERIOD OF PLATEAU	TASK LEARNING	STARTING TRIAL
	INITIAL	PLATEAU	TRIALS	TRIALS	
1.BASELINE					
AIRCRAFT HEADGEAR AND INTERCOM (AIC-25)-NO NOISE NO JAMMING	65%	98%	27	15	1
2.NOISE					
ADD COCKPIT NOISE AT 105dB- NO JAMMING	86%	94%(Initial)	10	12	43
		90%(Final)	33		
3.NOISE/JAMMING					
ADD JAMMER A + 3dB J/S RATIO	10%	30%	22	61	99
4.NOISE JAMMING					
TWO SUBJECTS AFTER TWO WEEK BREAK	50%	40%	45*	--	183
FINAL TRIAL					228

* AVERAGE DATA FOR TWO SUBJECTS SHOWS GRADUAL
DECLINE IN PERFORMANCE FROM 50% TO 40% OVER
THE 45 TRIALS. THESE POST-BREAK DATA SHOW
WIDE VARIATION.

TABLE 3

The average initial response of the six listeners for the baseline communication condition was about 65% correct. Group performance improved rapidly over the initial 15 trials till it plateaued for 27 trials at 98% correct. Each trial required about 5 minutes to complete.

The simulated cockpit noise at a level of 105 dB was then introduced. Average intelligibility dropped about 12% initially. Performance again improved rapidly reaching 94% after 12 trials. Performance finally stabilized at about 90% correct, representing a loss of about 8% in intelligibility due to the presence of the noise.

Next the modulated noise jamming signal was added to the communication link in the presence of the simulated cockpit noise and 83 additional trials were run. The jamming signal severely degraded the word intelligibility resulting in an average score of 10% correct. This level of performance remained essentially constant for the next 27 trials, during the subsequent 35 trials group performance improved reaching a plateau at about 30% correct. It should be noted that this is approximately the same level of performance shown by the trained listener panel used in the jammer effectiveness study under comparable conditions. (See Fig 2).

Two listeners returned to the laboratory for additional measurements following a two-week break. Although the average score at the end of the experiment was 30% correct for all subjects, the two returning subjects had performed well above average with individual scores of 40% and 45% correct. Their post-break average performance was about 45% correct, actually varying from a little over 50% correct during the initial post-break trials to around 40% over the latter trials.

The training effect under investigation in this pilot study was evident for the naive listeners at each stage of the study where a new or different task was introduced. The listeners' initial responses to word intelligibility materials presented over a standard AF aircraft communication system exhibited progressively increasing performance over the first 15 trials.

The introduction of cockpit noise and later of an acoustic jamming signal each caused a reduction in performance. In both cases the listeners' performance progressively improved until a plateau was reached. Two subjects showed essentially the same level of performance after a two week break as they had displayed prior to the break. The results of this exploratory study indicate that a comprehensive, well designed investigation directed to a wider range of conditions and jammers and addressing individual subject differences is warranted. The results of such a study may aid in developing methods of effectively selecting and/or training individuals who must perform in the presence of jammed communications prior to their operational assignments.

REFERENCES

1. Horn, R. A. "Analysis of Combat Air Missions," MITRE MTR-3030, 1976.

2. McKinley, R. L. "Voice Communication Research and Evaluation System," AMRL-TR-80-25, 1980.

3. House, A. S., Williams, C. E., Hecker, M. H. L., Kryter, K. D. "Articulation Testing Methods: Consonantal Differentiation with a Closed Response Set," J. Acoust. Soc. Amer., 37, 1965, 158-166.

4. Webster, J. C. and Allen, C. R. "Speech Intelligibility in Naval Aircraft Radios," NELC-TR-1830, Aug 1972.

5. Ascher, J. W., Hanley, T. D. and Steer, M. D. "An Investigation of Listener Accuracy in an Environment of Relevant Conflicting Voice Signals," SPECDEVCEN-TR-104-2-39, Dec 1954 (AD639-104).

6. Nixon, C. W. McKinley, R. L. and Moore, T. J. "Increase in Intelligibility of Jammed Word Intelligibility Tests due to Training of Listeners," (in preparation).

Phonological Variants in Medial Stop Consonants under Simulated Operational
Environments: Implications for Voice Activated Controls in Aircraft

James D. Mosko, Ph.D., Roger W. Remington, Ph.D., and Glenn R. Griffin
Naval Aerospace Medical Research Laboratory
Naval Air Station, Pensacola, Florida 32508

SUMMARY

A motion disorientation test was used to determine potential effects on voice characteristics. Speech samples from 14 subjects were obtained under a STATIC (control) and a DYNAMIC (experimental)condition. The 10 subjects completing both phases of the test did not exhibit significant changes in fundamental frequency, word token duration, or voice level. Four subjects who voluntarily curtailed the DYNAMIC mode exhibited significant changes in the same acoustic characteristics. The implications of these results for automatic speech recognition are discussed.

INTRODUCTION

The integration of Automatic Speech Recognition (ASR) technology into training and operational systems in the military is progressing rapidly due to the accelerated advances in micro-electronics. However, the human performance elements required to implement ASR technology successfully have not been defined fully. Serious deficiencies exist in our knowledge of the expected variations in the phonological characteristics of human speech in the physiologically and psychologically demanding conditions of modern military operations. Knowledge of such variations will provide information necessary to produce ASR algorithms capable of operating in these military environments. (1, 2, 3, 4 and 5.)

The data presented here represent the results of an initial experiment of an extended project designed to provide detailed information on the phonological, lexical, grammatical, and syntactical characteristics of aircraft voice communications in routine and demanding operational environments. This investigation deals with the spectral and temporal variations in the acoustic realizations of the alveolar and palatal stops /t, d, k, g/ in a condition of motion disorientation, a condition not uncommon to military aircraft operations and one which is both physiologically and psychologically demanding. Tobias (6) has shown the combined frequency of occurrence of these phonemes to exceed 17% in American English. Preliminary data from another aspect of this project indicate a combined frequency of occurrence in excess of 25% in aircraft communications. Apparently, these phonemes carry an important information-bearing load in such situations.

For this investigation, the stops appeared in the medial position in each target word. "Medial" is defined as /t, d, k, g/ appearing between two vowels, or those with a sonorant between the stop and the preceding vowel. This definition was used in order to compare any acoustic variations encountered with the extensive data on medial stops presented by Zue and Laferriere (8).

METHOD

A brief list of 10 words was constructed to meet the constrained definition of "medial" and to reflect words commonly used in aircraft communications. All words were poly-syllabic and two randomized 50-item word lists were constructed to conform to the response paradigm of the Visual-Vestibular Interaction Test (VVIT) (7).

In the VVIT, the subject is placed in an aircraft seat facing a lighted matrix containing letters on the top margin and digits on the left margin. A coordinate cue is presented aurally to the subject whose response task is then to locate the intersection of the coordinates and to call out the digit appearing beneath the letter and to the right of the digit, and the next two digits directly under it in the column. A subject performs a STATIC and a DYNAMIC (sinusoidal oscillation, .02 Hz, 30 rpm peak) mode.

For this study, in addition to the aural coordinate cue, the subject received one of the target words aurally (following the coordinate cue) and his task was to call out the target word followed by the three digits. As in the VVIT, this task was performed in both the STATIC and DYNAMIC modes. The vocal responses of each subject were recorded on magnetic tape using a General Radio electret microphone circuited to a Nagra IV-SJ (Kudelski) recorder for subsequent analysis.

Analysis consisted of extracting the target tokens from the response phrases and acoustically analyzing these tokens using both a speech spectrograph (Kay, Model 7029A) and the Visi-Pitch (Kay, Model 6087). Answers to the following questions were sought through this analysis:

(1) How do the spectral and temporal characteristics of similar tokens differ between the STATIC and DYNAMIC modes?

(2) Are interspeaker differences present?

(3) In the limited context of the experiment, do the variations follow the phonological rules set forth by Zue and Laferriere (8)?

(4) What are implications of the variations for ASR technology?

RESULTS

Fourteen male volunteers served as subjects for this experiment: 11 U. S. Marine Corps Student Naval Flight Officers (SNFOs) and three of the experimenters connected with the study. Ten SNFOs subjects completed both the STATIC and DYNAMIC phases of the VVIT and four subjects only completed from one-third to one-half of the DYNAMIC phase. The presentation of the data has been partitioned to show results for the 10 subjects completing both phases of the VVIT and for the four subjects who completed only part of the DYNAMIC phase. For the subjects completing both phases, the data represents averages over five samples of each of the word tokens for each subject in both phases. For the subjects not completing the DYNAMIC phase, the data represents averages over five samples for each word token in the STATIC phase, but for only three samples in common, completed word tokens in the DYNAMIC phase.

The data in Figure 1 portray the changes in mean fundamental frequency (f_o) for each of the 10 subjects completing both phases. The mean f_0 for each subject in the STATIC mode is within the expected range for young adult males. Fundamental frequencies obtained for the DYNAMIC mode indicate the individual differences in reactions to the oscillation and performance task. Although a general trend appears for the f_0 to increase in the DYNAMIC mode, the increase is slight and the differences shown are not statistically significant. Two of the subjects exhibited a small decrease in the f_o.

The mean durations of the word tokens for each subject completing the two phases of the task are depicted in Figure 2. Here again, the differences noted between the two phases are not statistically significant and a trend toward lengthening or shortening the tokens cannot be noted.

A measurable increase in the level of voice is depicted in Figure 3 for those who completed both phases of the VVIT. The differences shown indicate an increasing voice level in the DYNAMIC phase. However, the noise associated with the VVIT (approximately 68 dBLp) can account in part for the increases exhibited. Therefore, the significance of the voice level increases in these subjects cannot be assessed with these data.

Data on the mean duration, f_o, and voice level for the four subjects who did not complete both phases of the VVIT are portrayed in Figures 4, 5, and 6. Since these data are based on limited samples for the DYNAMIC phase, the statistical significance of the data must be suspect.

The mean f_0 for each of the four subjects appears in Figure 4. In general, the f_0 for each subject changes significantly in the DYNAMIC mode. Three of the subjects show increases in the f_0, while one subject exhibited a small decrease in his f_o. The mean changes are larger than the changes indicated earlier for the subjects completing both modes.

The duration of the word tokens, as reflected in Figure 5, does not indicate major deviations between the two phases of the task. Lengthening of the duration appears as a trend in these data, as with the earlier data; but the changes are not large enough to assign significance to their occurrence.

Vocal output level variations between the two modes of the VVIT are shown in Figure 6. In contrast to the data presented earlier, the variations for the four subjects are larger and cannot be attributed totally as a reaction to the noise generated by the power source for the oscillations.

DISCUSSION

The scope of the data precludes drawing definite conclusions regarding the expected values of voice characteristics in the experimental environment. The nature of the changes suggested by the data does follow predictable variations when one considers possible strategies for coping with motion disorientation.

The change in f_o noted for the subjects, particularly, the four who did not complete the DYNAMIC mode, can reflect one of two techniques for coping with motion stress--increased subglottal pressure and/or increased tension of the laryngeal musculature. If the respiratory patterns are changed, an increase in the subglottal pressure can result in an increase in the f_o. Similarly, tensing of the glottal folds can result in an increased f_o. An opposite effect (lowering of the f_o) occurred in three of the subjects. This can reflect an attempt to relax and breathe deeply in order to cope with an onset of an uneasy feeling. An increase of tension can also be reflected in resonance changes in the vowels or changes in the articulatory patterns while speaking. The data derived from this initial experiment do not indicate changes of this nature.

The increases in the f_o of the subjects are confounded by the increase in the vocal output. An increase in the f_0 can be expected to parallel a significant increase in the vocal level. The trend toward the occurrence of both in the current data, do not permit speculation regarding the major contributing factor toward the occurrence of either effect.

A lack of change in the durations or phonological variations for the word tokens indicates, even for the subjects who curtailed the DYNAMIC mode, the effects of the task were not sufficient to impart major changes in the articulatory patterns. Thus salivation or dryness of the mouth, both of which are possible in individual coping strategies, apparently did not occur for the current task.

Implications for ASR are present in the current data. Changes in f_0 and vocal level can be expected to occur in individuals, under conditions of motion stress. These changes will, in general, be outside the envelope of the normal variations expected to occur in ordinary operational environments. Thus, the integrity of an operational ASR device must take into account the variations when they occur. Additional research along the lines presented here will provide the information necessary to define the expected changes.

REFERENCES

1. Beck, B., Newberg, E.P., and Hodge, D.C. An assessment of the technology of automatic speech recognition for military applications. IEEE Trans. Acoust. Speech Signal Processing, Vol. ASSP-25, No. 4, pp. 310-322, August 1977.

2. Cannings, R., Borland, R. G., Hill, L. E., and Nicholson, A.N. Pitch and formant analysis of the voice in the investigation of pilot workload. AGARD Conference Proceedings, No. 216, Methods to Assess Workload, AS-1-45-11, 1978.

3. Fuege, R. L. and Geer, C. W. Integrated applications of automated speech technology. (Final Report. ONR Contract N00014-77-C-0401). Boeing Aerospace Company (with Logicon, Inc.), February 1978.

4. Lea, W. A. and Shoup, J. E. Gaps in the technology of speech understanding. Paper presented at the International Conference of Acoustics, Speech, and Signal Processing, Tulsa, Oklahoma, 1978.

5. Strieb, M.I. and Fisher, C. The voice recognition and synthesis (VRAS) system. (Analytics Technical Report #1234). NADC Contract #N62269-77-C-0158, 1977.

6. Tobias, J. V. Relative occurrence of phonemes in American English. J. Acoust. Soc. Amer., Vol. 31, p. 631, 1959.

7. Lentz, J. M. and Guedry, F. E. Motion sickness susceptibility: A retrospective comparison of two laboratory tests. Aviat. Space and Environ. Med., Vol. 49, pp. 1281-1288, 1978.

8. Zue, V. W. and Laferriere, M. Acoustic study of medial /t,d/ in American English. J. Acoust. Soc. Amer., Vol. 66, No. 4, pp. 1039-1050, October 1979.
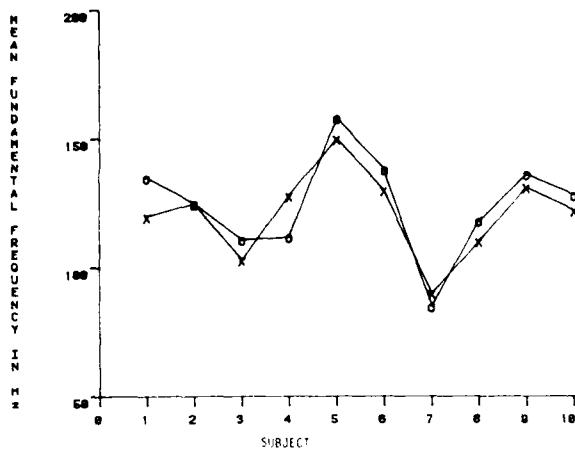
Figure 1. Mean fundamental frequency (f0) for 10 subjects in the STATIC (x) and DYNAMIC (o) phases of the VVIT.
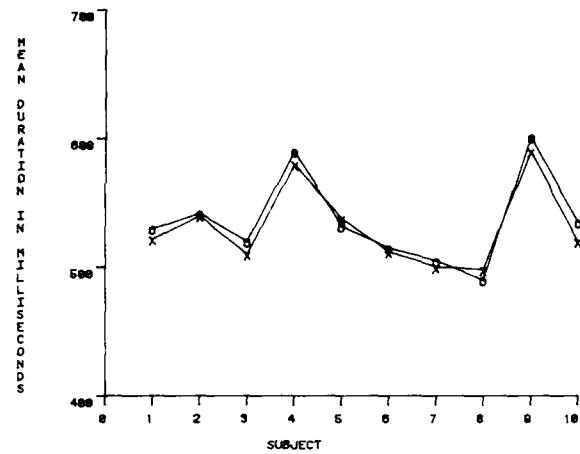
Figure 2. Mean word token durations (in msec) for 10 subjects in the STATIC (x) and DYNAMIC (o) phases of the VVIT.
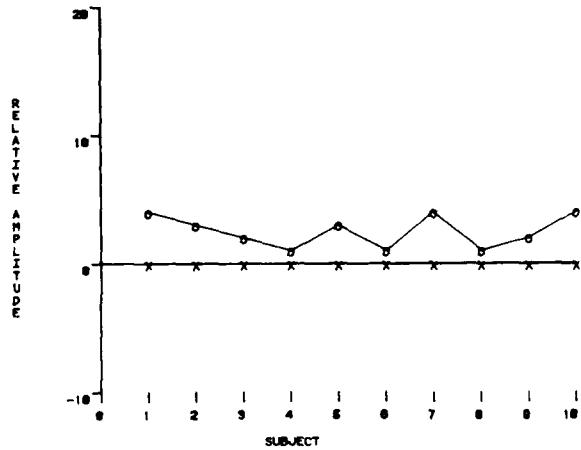
Figure 3. Relative amplitudes of voice levels (normalized to the level in the STATIC mode) obtained in the DYNAMIC mode of the VVIT.
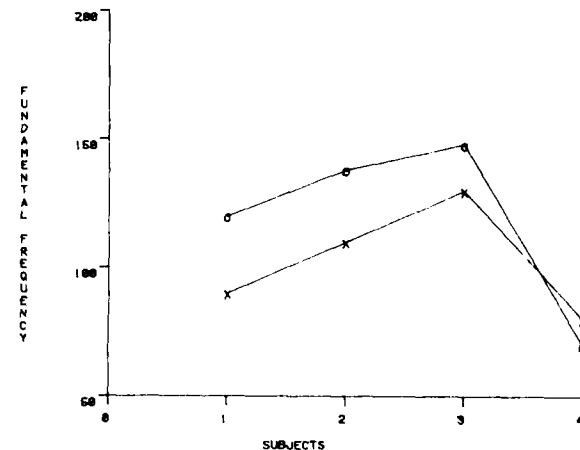
Figure 4. Mean fundamental frequency for the two modes of the VVIT (x = STATIC, o = DYNAMIC) for four subjects not completing the DYNAMIC mode.
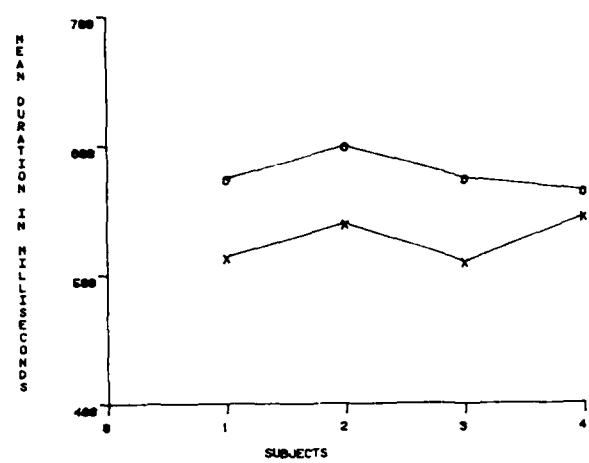
Figure 5. Average word token durations for four subjects not completing the DYNAMIC phase of the VVIT. (X = STATIC, O = DYNAMIC)
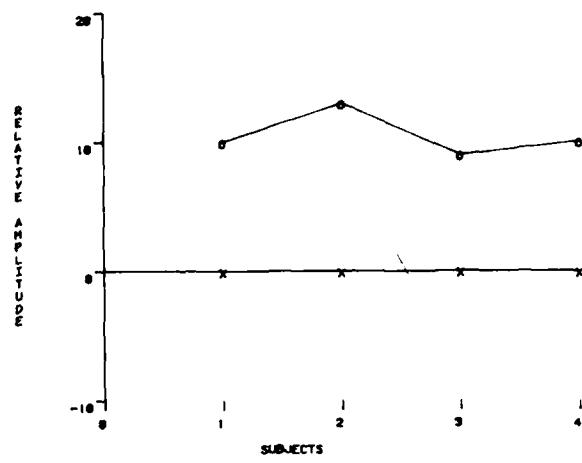
Figure 6. Normalized relative amplitudes of voice levels for four subjects not completing the DYNAMIC phase of the VVIT. (X = STATIC, O = DYNAMIC.)

"Clear Speech": A Strategem for Improving Radio Communications and
Automatic Speech Recognition in Noise

James D. Mosko, Ph.D.
Naval Aerospace Medical Research Laboratory
Naval Air Station, Pensacola, Florida   32508

SUMMARY

The acoustic characteristics of Conversational Speech production and Clear Speech production were compared for three different talkers. Increases in fundamental frequency, word token duration, and voice level for the Clear Speech were obtained. These results are compared to the results of similar studies and implications for improved intelligibility of speech and automatic speech recognition are discussed.

INTRODUCTION

The intelligibility of given message depends on the manner in which the word or phrase tokens are spoken; the size of the lexicon from which the tokens are extracted; the "noise" in the channel of communication; and the individual variations which occur in any of these. With all other factors held constant, the variations in token production, both within and between speakers, has been of continuing interest to communication engineers and speech scientists. The variations in the intelligibility of speakers assumes additional importance with the current trend toward incorporating speech recognition devices for query and control functions in aircraft and digital voice transmission technology for radio communications (in particular, low bit rate digital devices.)

Recently, Picheny and others (1, 2, 3) at the Massachusetts Institute of Technology reported data concerning a "clear speech" method of speech production, the resultant improvement in intelligibility scores of hearing-impaired listeners, and the acoustic characteristics of the "clear speech" tokens which might contribute to changes in the intelligibility. "Clear speech" was defined as verbal word or phrase tokens spoken with the talker specifically instructed to articulate clearly. This technique can be contrasted to "conversational speech" in which the talker produces the tokens in a manner consistent with his usual production patterns.

Picheny and Durlach (1) compared the intelligibility scores obtained by four hearing-impaired listeners when sentences produced by a single talker were presented to them in both the "clear speech" and "conversational speech" modes. All four listeners exhibited higher intelligibility scores--18 percentage points, on the average -- with the "clear speech" tokens. Similarly, Picheny, Durlach, and Braida (2) reported comparable improvements in intelligibility scores for hearing-impaired listeners using multiple speakers and syntactically normal but semantically ambiguous sentences.

These data prompted speculation regarding: (a) whether the "clear speech" technique would improve the intelligibility of radio voice communications; (b) what acoustic changes in speech tokens occur as a function of "clear speech" production; and (c) what effect these changes would have on the performance of automatic speech recognition devices. A series of experiments was designed to address these issues. This paper presents preliminary data from one experiment.

METHOD

Speakers. Three speakers, two males and one female, were trained to produce speech tokens in a "clear speech" mode. Following explicit instructions by the investigator on the manner in which they were to "overarticulate" the speech tokens, each speaker practiced--in the presence of the investigator-- approximately four hours on the "clear speech" production mode. When both the investigator and each speaker were satisfied with the mastery of the technique, each speaker then recorded the experimental materials. The experimenter served as one of the speakers.

Speech-token materials. The speech tokens were chosen to include both words and digit sequences and to reflect typical voice transmissions to and from aircraft. The digit sequences were derived from the response series used in the Visual-Vestibular Interaction Test (4) for the purpose of acoustical comparisons with data obtained in a parallel experiment. The word tokens were chosen based on two criteria: (1) the tokens occur commonly in aircraft communications, and (2) stop consonants occur in the medial position of each word. Each word token occurred randomly at least four times in each of two 50-item lists. The same trailing digit sequence did not occur with each word. Each spoken phrase took the form: "Say," "Target word," "digit sequence."

Recording paradigm. In order not to contaminate the conversational speech mode with the practice of the "clear speech" mode, the conversational mode was always recorded first. A speaker practiced producing both the word tokens and digit sequences separately and in the experimental form until complete familiarity with the tokens was accomplished. The speaker then provided 10 consecutive samples of each word token and digit sequence, spoken separately. The speech samples were recorded in this manner to be used subsequently to "train" an automatic speech recognition device.

All of the recording sessions were conducted with the talker seated in an acoustically isolated room. The chair for the speaker was fitted with a headrest in order to maintain a constant mouth-to-microphone distance of 30.6 centimeters and normal incidence. The Bruel and Kjaer Type 4135 condenser microphone, on an adjustable stand, input the token samples to a Nagra IV-SJ (Kudelski) magnetic tape recorder located outside the room. When the speaker was producing tokens in a noise background, the noise was produced by a General Radio generator (Model 1382) amplified, attenuated, and presented to the speakers via TDH-39 earphones fitted with supra-aural cushions. The noise levels, 70 dBLp and 90 dBLp (re: 20μPascals), were

measured with a General Radio sound level meter (Type 1565-A) associated with a General Radio 9A coupler (Type 1560). Calibration of both the recording apparatus and the noise generating apparatus was conducted routinely before and after each recording session.

Following the obtaining of the conversational training samples, the speaker produced conversational samples of both experimental lists in a no-noise background condition, a background noise condition with a level of 70 dB at the ears, and a background noise condition with an 90 dB level at the ears. These three conditions were randomly assigned to each talker. "Training" samples also were obtained for each of the noise conditions. The recording conditions were the same for the clear-speech production mode.

Analyses. For purposes of the present study, acoustical comparisons of the clear-speech and conversational speech tokens were made to determine variations in (1) length of token utterances; (2) relative intensity of token utterances, (3) fundamental frequency of token utterances; and (4) consonant-vowel (C/V) ratios within the utterances.

RESULTS

Figures 1 through 3 portray the relevant fundamental frequency (f_o), relative amplitude, and token duration data obtained for the Conversational Speech production mode.

The f_o for each subject displayed in Figure 1 appears relatively unchanged for the Quiet and 70 dB background noise conditions. For the 90 dB background noise condition, however, each subject's f_o increased 26 Hz on the average. The increase in f_o is not unexpected for the 90 dB noise case, since a significant increase in the amplitude of the voice can be expected in noise of this magnitude.

Relative amplitude increases for the three recording conditions appear in Figure 2. Very small (3 dB on the average) changes in voice level can be noted when the background noise level is 70 dB. In the 90 dB noise case, voice level increases on the order of 10 dB appear for each subject and are reflected in the parallel increases in f_o.

Mean durations of the word tokens for each subject are depicted in Figure 3. The lack of durational differences between the three conditions is somewhat surprising, since one might expect some changes, particularly in the 90 dB background noise condition.

Measurement data for the Clear Speech production conditions appear in Figures 4, 5 and 6. In Figure 4, the range of the f_o for each speaker is similar to the range for the Conversational Speech mode. Two interesting occurrences can be noted in this data. The f_o for Speakers 1 and 2 is increased significantly, even in the Quiet condition. However, Speaker 3 exhibits a decrease in the f_o, although the range for the three conditions is similar to the Conversational Speech case (about 22 Hz).

Voice levels for the Clear Speech condition, shown in Figure 5, also exhibit changes similar to those in the comparable data for the Conversational Speech. This result is somewhat unexpected and will be discussed below.

Major lengthening of the word tokens in the Clear Speech case appears in the data of Figure 6. Subjects 1 and 2 exhibit some relative decrease in word token duration for the noise conditions, but Speaker 3 does not.

Differences in token durations between the Conversational mode and the Clear Speech Mode are scaled in Figure 7. All subjects exhibited significant lengthening of the word tokens, especially Speaker 1. Although some lengthening was to be expected, the data for Speaker 1 was excessive. A comparison of the voice levels between the Conversational and Clear Speech conditions (Quiet) is portrayed in Figure 8. A 14 dB increase in voice level in the Conversational Speech should be noted. The voice level increase is comparable to the increase for the Conversational Speech, 90 dB background noise condition. This increase can account for the increases in f_o for Speakers 1 and 2 shown in Figure 4.

The relative amplitudes between the consonants and vowel (C/V ratio) were measured for the word tokens where possible. Although not shown in graphical form, some changes in the C/V ratios were obtained. In general, the speakers in this study exhibited larger C/V ratios (for example, -6 to -10 dB) for initial stressed syllables in the tokens and smaller C/V ratios (for example, -8 to -2 dB) for unstressed or secondary stress syllables.

DISCUSSION

The data presented here correspond to the changes in f_o, token duration, relative amplitude, and C/V ratio reported by Picheny and others, particularly in the direction of the changes. The present data, however show increases which are in excess of those reported by Picheny (1). For example, Picheny reports a 6-8 dB increase in the voice level for the Clear Speech mode, while the data from this experiment indicate increases of 10 dB. Similarly, Picheny reports duration increases on the order of 0.5 to 1.6 for Clear Speech. Two of the speakers in this experiment (Speakers 2 and 3) exhibited increases in line with Picheny's speakers, but Speaker 1 is well above expected values. Variations between the results of the two studies can be attributed to individual speaker variations and/or differences in the instructions given to the speakers. Additional research will clarify the reasons for the discrepancies.

Other unexpected results appeared in the data from this study. For example, the increases in the voice level for Clear Speech in the background noise conditions are considered unusual, if one notes the "natural" tendency of the speakers to increase voice level in the Clear Speech, Quiet Condition. One might expect the increase to override any influence of the background noise to hamper the voice monitoring abilities of the subjects. Apparently, for the speakers in this experiment, this was not the case, and voice level increases similarly to those obtained in Conversational Speech.

Clear Speech can result in a 16-18 percent improvement in intelligibility scores in quiet listening conditions (1). Formal listening tests have not been completed with the word tokens derived in this experiment, but some generalizations can be made from informal tests. At a signal-to-noise (S/N) ratio of 0 dB, two listeners improved their speech intelligibility scores by six to eight percent for one speaker (Speaker 3). This increase is not on the order reported by Picheny (1), but the occurrence of the increase supports the idea of using Clear Speech techniques for adverse S/N ratios and the idea of placing more emphasis on "training" radio operators to optimize articulatory gestures when attempting to communicate, especially in cases where the message sets are limited, as in aviation environments.

The potential increase in S/N ratio and the improvement in C/V ratio during Clear Speech production points strongly to improvements which can be expected using this technique.

REFERENCES

1. Picheny, M.A. and Durlach, N.I. Speaking clearly for the hard of hearing. J. Acoust. Soc. Amer., Suppl 1., Vol. 65, S135(A), 1979.

2. Picheny, M.A., Durlach, N.I., and Braida, L.D. Speaking clearly: Intelligibility and acoustic characteristics of sentences. J. Acoust. Soc. Amer., Suppl 1, Vol. 67, S38(A), 1980.

3. Chen, F.R., Zue, V. W. , Picheny, M.A., Durlach, N.I., and Braida, L.D., Speaking clearly: Acoustic characteristics and intelligibility of stop consonants. J. Acoust. Soc. Amer., Suppl 1., Vol. 67, S38(A), 1980.

4. Lentz, J. M. and Guedry, F.E. Motion sickness susceptibility: A retrospective comparison of two laboratory tests. Aviat. Space and Environ. Med., Vol. 49, pp. 1281-1288, 1978.
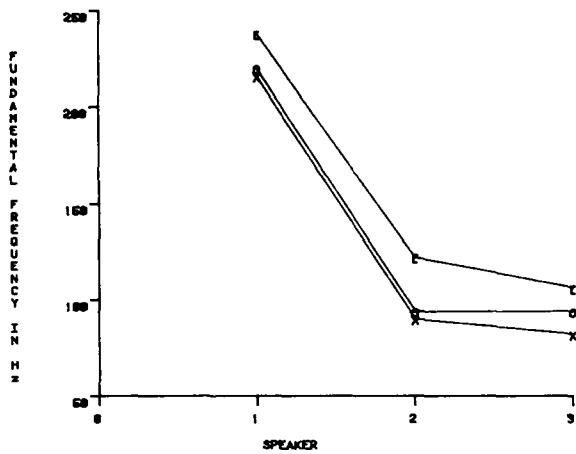
Figure 1. Average fundamental frequency (f_o) for the three speakers in the Conversational Speech production mode, for three speaking conditions (X = Quiet; 0 = 70 dB background noise; [= 90 dB background noise).
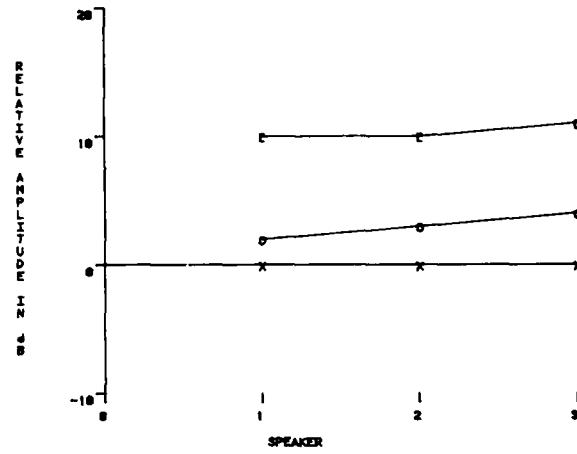
Figure 2. Relative amplitudes of voice levels for three speakers in the Conversational Speech mode, normalized to the level obtained in the Quiet speaking condition (X = Quiet; 0 = 70 dB background noise; [= 90 dB background noise).
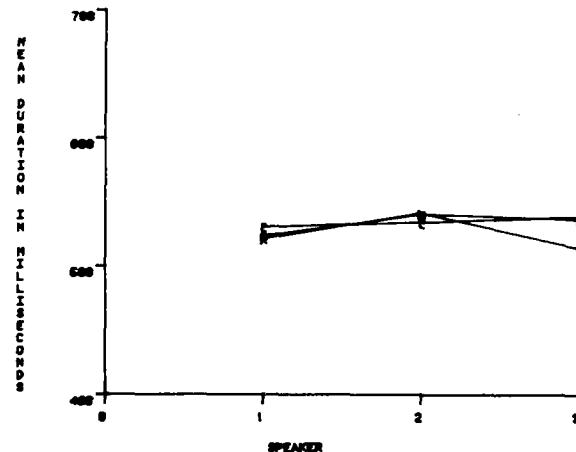
Figure 3. Average word token durations for three speakers in three speaking conditions (X = Quiet; 0 = 70 dB background noise; [= 90 dB background noise)
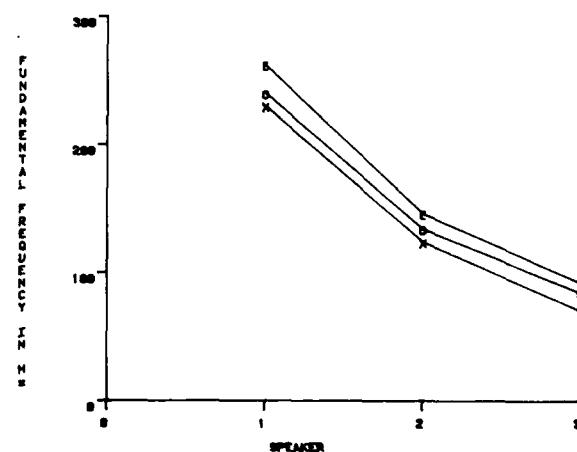
Figure 4. Fundamental frequencies for three speakers in Clear Speech production mode, for three speaking conditions (X = Quiet; 0 = 70 dB background noise; [= 90 dB background noise).
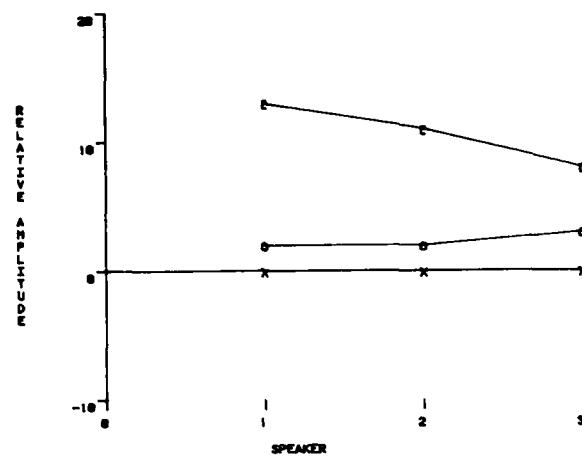
Figure 5. Voice levels for the Clear Speech production mode. (X = Quiet; 0 = 70 dB background noise; C = 90 dB; background noise)
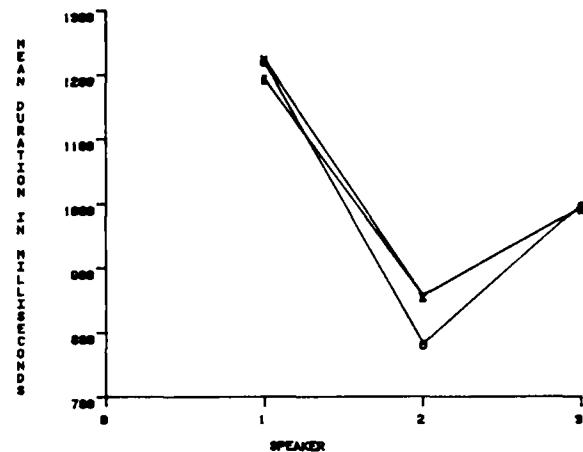
Figure 6. Word token durations in the Clear Speech production mode (X = Quiet; 0 = 70 dB; C = 90 dB).
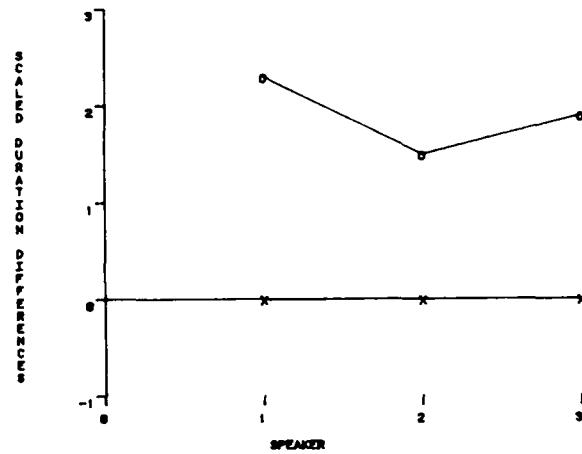
Figure 7. Durational differences (expressed as ratios) for the three speakers in the Quiet condition, for Conversational and Clear Speech production mode (ratio is normalized to the Conversational Speech mode). (X = Conversational; 0 = Clear)
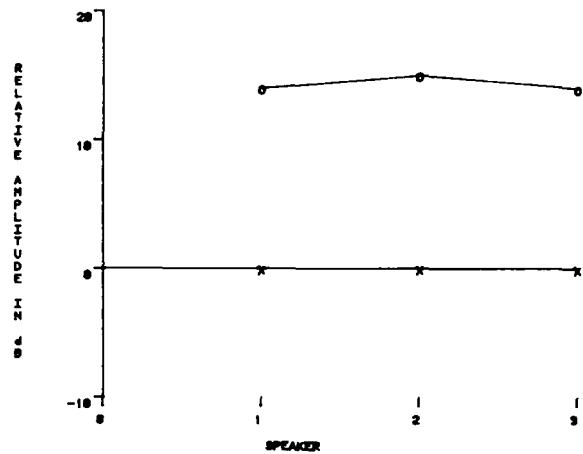
Figure 8. The contrast between voice levels for the Conversational and Clear Speech modes (X = Conversational, 0 = Clear).

DISCUSSION

QUESTION

I did not get a clear impression of what you meant by clear speech. How did you train your subjects?

DR. J.D. MOSKO (UNITED STATES)

Our instructions to the subjects were simply try to over-articulate, STET for example, all of the syllables in a word without changing, or changing as little as possible, your natural timing and so forth, in speech. For example, they were trained for approximately 3-4 days before I was satisfied that they were articulating the way that I would like. I made all of my clear speech recordings after my conversational speech recordings so that I would not contaminate the conversational speech. It was simply a matter of listening to them produce these over and over again (i.e., reading prose, reciting poetry) just to get them into the habit of producing speech this way and then they practice on the actual words that they were to use plus we had the digits and so forth.

QUESTION

I was intrigued by your title, especially Number 09, "Clear Speech: a Strategem for Improving Radio Communications and Automatic Recognition in Noise", especially the second part on Automatic Speech Recognition in Noise. Your very last sentence was, "For automatic recognition, clear speech is not very good because the duration is too long." Is the conclusion given in your title wrong?

DR. J.D. MOSKO (UNITED STATES)

Yes, it is wrong. You will find at least an improvement in the actual speech communications with humans. The tests that I have made with my automatic speech recognition device, with the clear speech in noise, have not been very fruitful. What I did for example, I trained in the clear speech in noise and then ran the samples back through it using clear, conversational speech and got about 85-90% correct recognition by the computer. I did the same training with the device using the clear speech and even with all of the samples that I had, I only got about 60% correct recognition by the computer.

QUESTION

If you had tested the procedure with an actual system, I wonder if by testing it did you also use clear speech as a reference itself, for the system?

DR. J.D. MOSKO (UNITED STATES)

Yes, I did. I cross-trained under the conditions.

QUESTION

Even then your score was only 60% correct?

DR. J.D. MOSKO (UNITED STATES)

Apparently, the individual samples by cross-training with this device you need at least 10 consecutive inputs of a particular word STET token to train the device. So I trained the device, then I played the clear speech samples through these, the ones they did randomly. Apparently, the variations within the individual clear speech tokens were sufficiently large to really keep the system from operating optimally. I do not have any good explanation for it right now. We're trying with the other levels of speech, the 70 dB and 90 dB, to see if this holds true.

QUESTION

1. What speech recognition system or program were you using?
2. Did it contain dynamic programming in order to account for time differences?

DR. J.D. MOSKO (UNITED STATES)

1. The Interstate Voice Recognition Module.
2. Yes, there is dynamic programming.

DR. R.L. MCKINLEY (UNITED STATES)

Your references to standard language has prompted me to make these remarks about something I have found in examining tapes of last messages from pilots during accidents. It is that usually, the message is a short unfamiliar language and in many cases, unintelligible. I think they could have been intelligible if the system had been designed correctly. As long as we are thinking about the standardization of language, we should keep this in mind, that we have the pigmelion effect of regression to a non-standard.

DR. J.D. MOSKO (UNITED STATES)

I agree. I think the situation where you have an unexpected event, perhaps we might over-train some of our talkers to use the system. A good example, as you are aware, if you take an air traffic controller who has just left school, and you listen to them, they speak very clearly, very concisely, very properly. You can almost chart how long they have been on the job by the deterioration in their speech and you notice this time and time again. When you train people to use radios - let's train them to use radios - they should be professional talkers.

DR. R.L. MCKINLEY (UNITED STATES)

I am not trying to discourage the need for training, but what I am trying to do is not let a special radio speaker make up for the quality of the system.

DR. J.D. MOSKO (UNITED STATES)

You need quality systems and quality speakers.

ELECTRONIC VOICE COMMUNICATIONS IMPROVEMENTS FOR ARMY AIRCRAFT

MITCHELL S. MAYER
AND
ARTHUR W. LINDBURG

US ARMY AVIONICS RESEARCH AND DEVELOPMENT ACTIVITY
FORT MONMOUTH, N.J.   07703

SUMMARY

The communications systems on all Army aircraft flying today are based on design concepts that are over 50 years old. The wide use of positive - peak-clipping (which introduces up to 50% distortion), limited bandwidths (300-3000 Hz), and unsophisticated AGC circuits reduce speech intelligibility and impact severely on aviator hearing loss. Outmoded test procedures, such as limited volume Kruff boxes for testing pressure gradient (noise cancelling) microphones and 6cc couplers for testing earphone elements, used in large volume circumaural earcups, remain in use throughout all branches of the service. These concepts and procedures are used because they have been standardized during the early years of communications electronics, and after many years of use, were accepted as sacrosanct.

After eight years of research and development, we have designed a totally modern, state-of-the-art communications system for Army aircraft, and have published two new specifications which contain many of the modern test procedures required to accurately test and evaluate the various components of the communication system.

As a first step in the development of new test procedures, we evaluated both ASA and ANSI standards and found them lacking. Their procedures for measuring the noise attenuation of hearing protective devices totally ignored the effects of the aircraft noise environment on microphone noise cancellation and earcup pumping; and the addition of necessary speech communications in the earcup, which must be at least 6-10 dB above the noise level in the earcup for adequate speech intelligibility.

The components of the new state-of-the-art communications system will include, as a minimum: high impedance DC powered noise cancelling microphones (using piezoelectric ceramic, electret, or PVF_2 diaphragms); earphone elements designed and tested to have flat frequency response when inside the circumaural earcup of the hearing protective device; and intercoms which replace positive peak-clipping with fast-acting AGC circuits and "expander/compander" circuits for maximum output signal without distortion, even under conditions of extreme stress.

In the future, audio signals in the microphone will be converted into the digital mode or directly into the optical spectrum for high efficiency, secure communications inside the aircraft. The savings in weight and security improvements will be considerable. Digitizing the speech signals will require research into minimum bit rates necessary for required speech intelligibility and time-sharing requirements when interfacing into the digital data bus (MIL-STD-1553B). Optical communications will require light-weight, efficient audio-to-optical or digital-to-optical converters. These converters along with optical amplifiers, couplers and splitters will be required to meet the same MIL-STD's all other airborne audio equipment must meet, and be easy to install and repair.

1. INTRODUCTION

The future battle scenario is envisioned to be a medium-intensity environment fought on or near the ground. This means that our aircraft will be flying and fighting in an NOE (Nap of the Earth) environment, drastically increasing the aviator's already heavy workload. The fatiguing effects of High Noise and poor communications (at NOE altitudes, line-of-sight radios work very poorly and for only short ranges) will have a deleterious effect on the aviator's combat effectiveness; therefore, the ability to communicate effectively in this harsh environment is essential to the successful accomplishment of his combat mission.

Military aircraft are designed for maximum capabilities in areas of performance, armament, and endurance. Acoustic noise suppression considerations, such as sub-chassis isolation of engines and transmissions, noise damping structural panels, and helmholtz resonator traps, impact on performance (added weight), as well as cost. (Total acoustic protection for the aircrew is not allowed to exceed 1-1.5% of the aircraft gross weight.) With these severe restrictions on noise attenuation at the source, personal aircrew protection is the only other avenue of approach. This sole protection, therefore, must be provided by the electronics package of the aviator's flight helmet.

The two major elements of the helmet electronics package are the noise-canceling microphone and the earcup/transducer assembly. For years these elements have been designed and tested to meet standard requirements and test procedures which do not realisticaly or accurately reflect the conditions under which they must operate. Before new communications elements can be designed and developed to meet the challenges of the high noise environments

of military aircraft, there must be a clear understanding of these standards and their shortcomings.

2. BRIEF HISTORY OF CURRENT STANDARDS

a. 300-3000 Hz Bandwidth. In the 1920's the Army began to realize the importance of speech communications in the battlefield. Bell Laboratories (Research Facility of the Bell Telephone Company) was tasked to supply the Army with the design considerations that must be applied to their communications system, in order for a successful system to be developed. One of these major design considerations was the requirement of a 300 to 3000 Hz communications bandwidth for the electronic transmission of human speech.

It is unfortunate that the 300 to 3000 Hz limited bandwidth, which worked so well for the telephone, is not adequate when communication is attempted while the talker or the listener is in a high noise environment. (Speech intelligibility becomes severely degraded when the ambient noise environment of the user exceeds 75 dBA.) After the first communications systems were in use in the Army, however, their shortcomings became apparent. One stop-gap measure (still in use today) was the use of a phonetic alphabet. This allowed a low intelligibility system to operate effectively in the military environment of that time.

Today, more than 50 years later, we still use this limited bandwidth in our communications systems; however, the use of the phonetic alphabet is not acceptable because it is slow and cumbersome, seriously degrading critical information acquisition in the medium-intensity NOE battlefield environment. The aviator (operating in the high-noise environment of Army aircraft) must be able to communicate effectively the first time. He does not have the luxury of repeating messages or using phonetically spelled words.

b. Coupler Calibration of Earphones (ANSI Z24.9-1949). The use of the 6cc (or Mott) coupler is another example of continuing to follow an outmoded "standard." The US Army, as with the bandwidth question, went to Bell Laboratories for a method of testing the frequency response of their communications headsets. Bell Laboratories recommended the 6cc coupler. (A detailed test procedure incorporating the 6cc coupler became ANSI Z24.9-1949.)

The 6cc coupler was developed for use as a 100% production-line test of Western Electric (manufacturing arm of the Bell Telephone Company) telephone handsets, because the closed volume between the earphone element and the eardrum approximates 6 cubic centimeters.

With today's large volume circumaural earcups, the use of the 6cc coupler for testing the frequency response of the installed earphone elements is highly misleading. The earphone element itself can still be measured using the 6cc coupler; however, when the earphone is placed into position inside the circumaural earcup, its frequency response is radically changed. These changes in the earphone frequency response impact significantly on speech intelligibility and cannot be seen when measuring the earphone as a separate entity in a 6cc coupler. A test procedure which measures the response characteristics of the complete earcup/transducer assembly is essential if realistic data is to be obtained.

c. Real-Ear Attenuation of Hearing Protective Devices (ASA STD 1-1975 and Z-24.22-1957, which it replaces). These two standards were excellent attempts to develop test procedures which would accurately test all types of hearing protective devices, including circumaural earcups. In most cases, these standards produce realistic results. The areas where these standards fall apart, however, are in high noise environments such as those that exist in military aircraft and when communications is included in the earcup. A hearing protective device that demonstrates acceptable attenuation values when tested by the threshold technique used in these standards, shows much lower values when tested in a high noise environment.

In a high noise environment, noise reaches the aviator's ears by several paths:

(1) Intense noise will penetrate the earcup directly, sending the entire assembly into sympathetic vibration. This "pumping" action causes the earcup to respond like a transducer, reproducing the noise inside the earcup.

(2) The "pumping" action of the assembly will cause the soft earcushion (resting against the head) to lift off the head, producing leaks in the seal, which allow noise to enter the earcup. Leaks in the earcushion seal can also be caused by improper helmet fit. An improper fit usually is the result of the user not fitting the helmet properly, or a user with a head size that does not fit properly into the two helmet sizes available to him.

(3) The noise cancelling microphone, when keyed, will detect the noise and pass it through to the communications system to be amplified and sent to the earphones in the aviator's helmet, as sidetone. It will also include this detected noise, along with the speech information, in the transmitted signal.

d. "Signal-to-Noise" Test Box (Kruff Box). While this procedure can compare two different noise canceling microphones and tell the tester that one may cancel noise better than the other, the data obtained using this procedure does not realistically reflect the operation of the microphone as it is used by an aviator. In actual use, the microphone is not inclosed in a limited volume, but is inside a large semi-reverberant chamber (the

aircraft), with one end of the microphone brushing against the aviator's lips and the other end surrounded by the noise field. The aviator's head acts as a partial block to the noise field, preventing the noise from impinging on both surfaces of the microphone equally, and thus severely limiting the noise canceling capability of the microphone. The Kruff box does not take into account the interaction of the human head and noise canceling microphone in an intense semi-reverberant noise field.

 e. <u>Positive-Peak-Clipping.</u> Another standard that has been in use for 50 years is the concept of positive-peak-clipping. When first introduced, it was thought that a square-wave audio output would provide the maximum signal strength to human speech, increasing the speech intelligibility of the transmitted voice to acceptable levels for radio communication. This technique is based on the fact that, while constants provide the major portion of sounds required for speech intelligibility, vowel sounds, being louder, mask the consonants in a sinusoidal signal. Using positive-peak-clipping, approximately 20% of the upper portion of the sinusoidal signal is clipped; however, the consonant sounds are now brought up to a level required for high speech intelligibility. (This technique, coupled with a fast rise and slow decay AGC circuit, is still in wide use today.)

 The major problem with positive-peak-clipping is that, when an amplifier goes into clipping, distortion of the audio signal can increase to levels approaching 50%. These very high levels of distortion decrease speech intelligibility so severely, that almost all the gains of increased consonant signal levels are lost. In some cases, this distortion can be so severe as to almost completely destroy all the speech intelligence. This condition exists when the aviator transmits a "panic message"; the one time that high speech intelligibility is essential.

 Today there are clamping circuits, such as "expander/compander" circuitry, that can give maximum signal strength, while still retaining a sinusoidal output even under "panic" situations. By expanding the frequency response of the communication system from 300-3000 Hz to 300-4500 Hz, the lost consonant sounds can be brought back into the transmitted signal without the need to positive-peak-clip the audio signal.

3. NEWLY DEVELOPED (REALISTIC) TEST PROCEDURES

 Now that an understanding of the shortcomings of current standards has been presented, a discussion of newly developed standards and test procedures can commence.

 a. <u>Increased Audio Bandwidth (300 to 4500 Hz).</u> High speech intelligibility in the military aircraft noise environment requires that all essential parts of the human speech spectrum (for both male and female voices) be reproduced and processed through the communication system. Since positive-peak-clipping has been eliminated in our new communication system design, this extended bandwidth is required to process the consonants in human speech which provide the speech intelligibility necessary in high noise environments. (Most consonants occur between 3000 and 4500 Hz.)

 In addition to a broader bandwidth for the communications system, the bandpass of the entire system should be relatively flat. This flat response, especially at frequencies below 1 KHz, increases speaker recognition which has proved to be an aid to increasing speech intelligibility in high noise environments.

 b. <u>Real Head Attenuation of Hearing Protective Devices in Pink Noise.</u> This procedure adds the requirement of a pink noise environment to the current standards for testing hearing protective devices. The sound pressure level of the pink noise in the test chamber should approximate the sound pressure levels experienced by aircrews during the performance of their missions. Two microphones are then used for measuring both the "ambient" noise environment and the "attenuated" environment at the ear. The use of a condenser microphone in the ambient environment and a miniature electret condenser placed on the conchea of the ear, protected by the circumaural earcup, are effective in obtaining the measurements required. If a 2-channel real-time analyzer is available, a real-time noise attenuation chart of the hearing protective device can be produced. The effects of helmet fit and movement inside the noise environment can then be evaluated, in real-time and recorded on a time histogram plot.

 The charts in Figures 1, 2, and 5 of this report were obtained from data supplied by a Two-Channel FFT (Fast Fourier Transform) Real-Time Analyzer (Spectral Dynamics Model SD-360). The same test subject wore both a standard SPH-4 aviator's helmet and a modified SPH-4 which contained a prototype MK-1564( )/AIC Headset-Microphone Kit. A 1/2 inch B&K condenser microphone was used to measure the ambient environment (Pink Noise at 105 dB SPL in the anechoic chamber and CH-47 helicopter noise at 115 dBA SPL in the aircraft noise environment simulator) and was connected to Channel A of the analyzer. A miniature electret condenser microphone was used to measure the "attenuated" noise at the subject's ear and was connected to Channel B of the analyzer. The analyzer then performed the transfer function: B/A. The data was plotted on linear paper and then transferred to semilogarithmic paper for this report.

 Figure 1 shows the differences in the attenuation characteristics of the US Army's SPH-4 aviator's helmet when measured according to ASA STD 1-1975 and when measured by the Real-Head procedure in pink noise at 105 dB SPL (average sound pressure level of US Army aircraft). (Note that the ASA standard procedure shows helmet attenuation at low frequencies to be greater than actually experienced in an intense noise environment. This

SPH-4 Helmet Attenuation - ASA STD 1-1975 Verses Real Head in Pink Noise
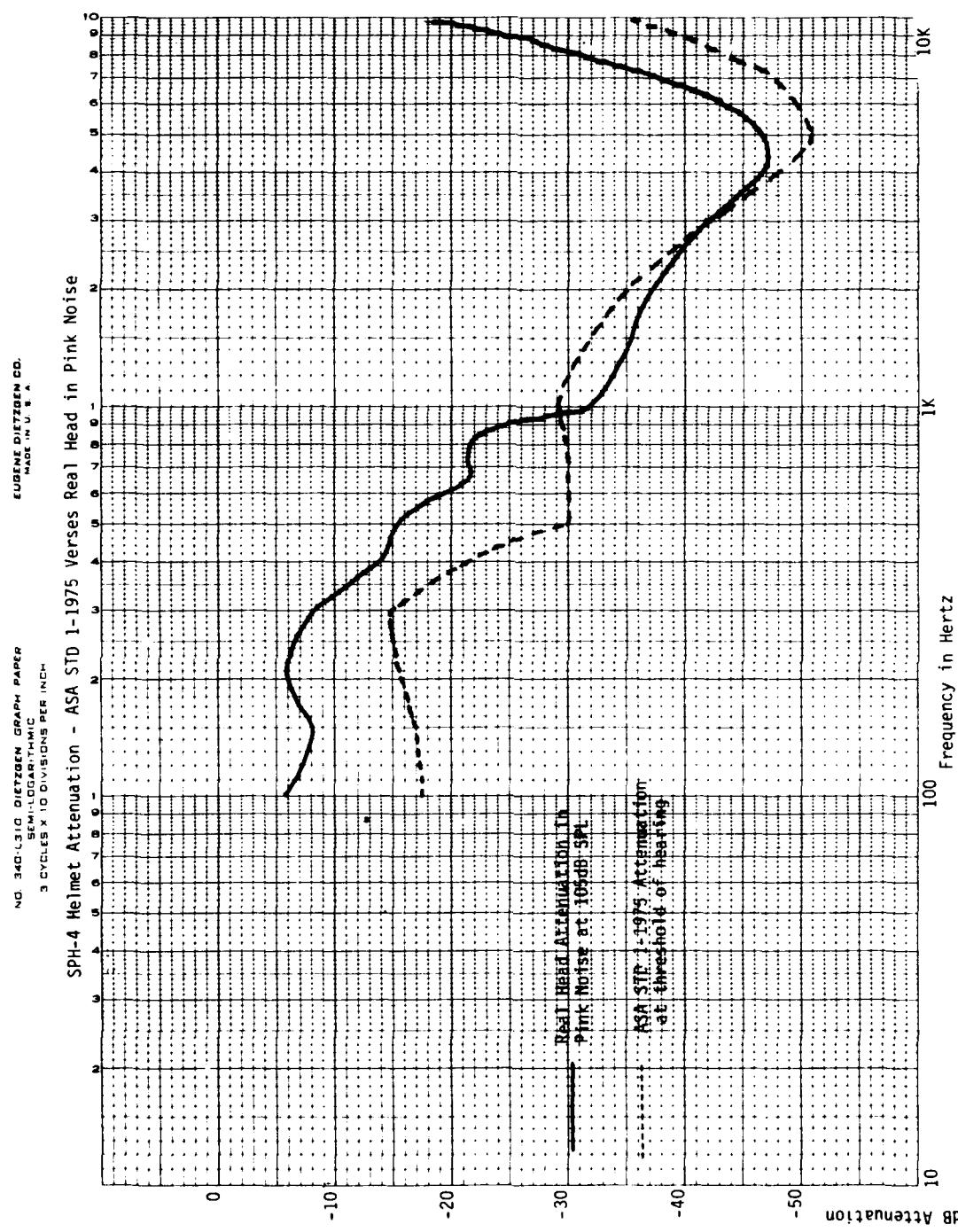
FIGURE 1

difference is mainly due to the "pumping" action of the earcups which can not be seen in the low level noise environments.)

Figure 2 compares the "real-head" in pink noise procedure to a "real-head" procedure in a simulated CH-47 Helicopter noise environment. (Note that these procedures compare favorably. There are differences of high frequencies because the CH-47 noise environment is at 115 dBA SPL compared to the pink noise at 105 dB SPL.)

c. **Microphone Noise Cancellation Procedure.** Microphone noise cancellation procedure requires two measurements, nearfield and farfield. The nearfield measurements are made using an artificial voice (Bruel and Kjaer Type 4219, or equivalent), while the farfield measurements require the use of an 8-inch speaker (normalized to a flat response, preferably by a digital computer) inside an anechoic chamber. The microphone under test must be at least one meter away from the sound source to be in its "farfield."

Noise cancelling microphones generally exhibit a cardioid polar response. To evaluate a microphone under "worst-case" conditions, the farfield response should be measured with the microphone facing the sound source. This measurement should not be averaged with measurements taken at the microphone's null points (90° or 270° incidence). This second method of measurement averaging shows greater noise cancellation for the microphone; however, the first method more closely represents the microphone's actual capabilities in a high noise environment.

d. **"Expander/Compander" Amplifier Circuitry.** The microphone amplifiers contained in the intercommunications controls should contain "expander/compander" circuitry to give maximum, distortion-free audio output to the transmitted signal. This circuitry should provide distortion-free sinusoidal output at the maximum voltage level necessary for proper transmitter modulation, regardless of whether the voice input is at a low level or at an extremely high level (as under emergency or panic situations).

4. COMPONENTS OF THE NEW AIRCRAFT COMMUNICATIONS SYSTEM

No one area of improvement previously mentioned, can increase speech intelligibility to an acceptable level (ideally 90% for high speech intelligibility). It takes improvements in all these areas, incorporated into a new communications system to achieve the necessary intelligibility levels. In the development of this system, old ideas have to be changed and new test procedures developed, to more accurately test and evaluate communications system components.

Three new specifications have been written to provide the US Army with a totally new communications system that will provide high speech intelligibility in the noise environments of military aircraft. These specifications are as follows:

MIL-C-49227(AV); Control, Communication System C-10414( )/ARC; 8 Sep 80.

MIL-M-49199 CR); Microphone, Linear, M-162/AIC; 30 May 80.

MIL-H-49198(AV); Headset-Microphone Kit MK-1564( )/AIC; 22 Oct 80.

(These specifications contain detailed descriptions of the new testing procedures briefly described in this paper.)

a. **The C-10414( )/ARC Intercommunication Control.** The C-10414( )/ARC is a combination microphone and headset amplifier which operates as a switchboard for each aviator. The C-10414( )/ARC, in that sense, is the same as its predecessors. The major improvements include the incorporation of "expander/compander" circuitry and a fast-acting AGC (automatic gain control) in the microphone amplifier; high isolation circuitry to eliminate cross-talk problems; audio limiting in the headset amplifier (this is necessary to prevent excessively high audio communications levels from damaging the aviator's hearing by adding to the existing high ambient noise and increasing hearing damage risk); and circuitry to power the FET (field-effect transistor) amplifier and impedance matching circuitry in the new M-162/AIC microphones. This intercommunication control, in conjunction with a new aircraft wiring harness which incorporates "balanced-line" techniques, has greatly improved the audio quality of speech communications, both inside the aircraft and that being transmitted out.

b. **The M-162/AIC Linear, High-Gain, DC Powered, Microphone.** The M-162/AIC incorporates a high-impedance voltage generating element (electret condenser, piezoelectric ceramic, or PVF$_2$) with an internal amplifier, to produce a flat nearfield frequency response (400 Hz to 6000 Hz ±3 dB). This microphone also provides a noise cancellation capability approximately twice as great as the currently fielded dynamic microphones. The nearfield and farfield frequency responses of a typical dynamic microphone (M-87) and a M-162 appear in Figures 3 and 4. (Note that the crossover point for the M-87 occurs at 1150 Hz, while for the M-162 it occurs at 2662.5 Hz. The area between the nearfield and farfield curves is the noise cancellation capability of the microphone. This capability is usually specified as the slope of the farfield curve starting at the crossover point and measured in dB/octave. For the M-87 microphone the average slope is: -2.9 dB/octave; for the M-162 it is: -6.3 dB/octave.)

c. **The MK-1564( )/AIC Headset-Microphone Kit.** The MK-1564( )/AIC will be the electronics package that will become part of the new product improved Army aviator's helmet, providing the improvements of the M-162 microphone and new linear earcup/transducer

Real Head Attenuation of SPH-4 Helmet in Pink Noise and Simulated CH-47 Noise
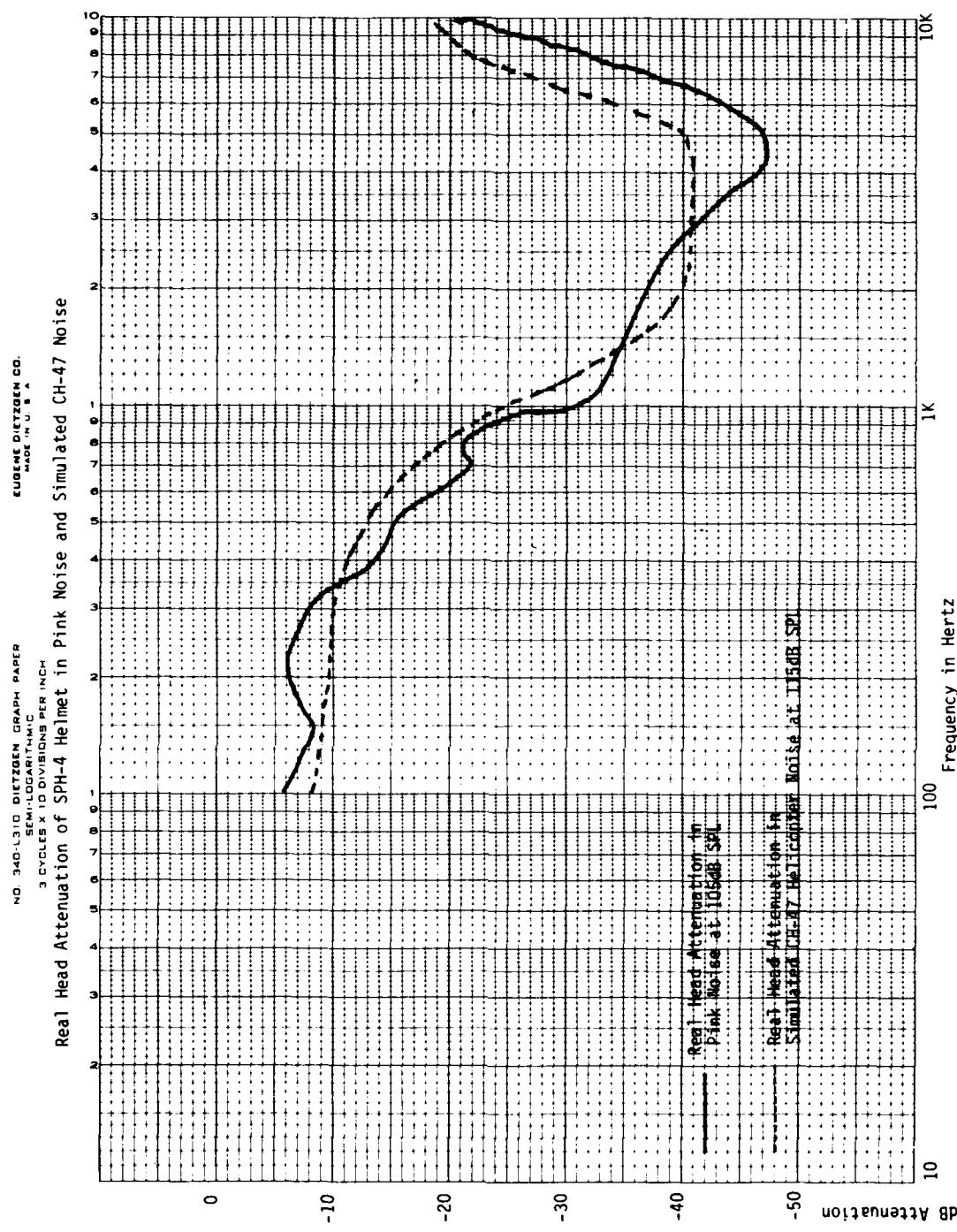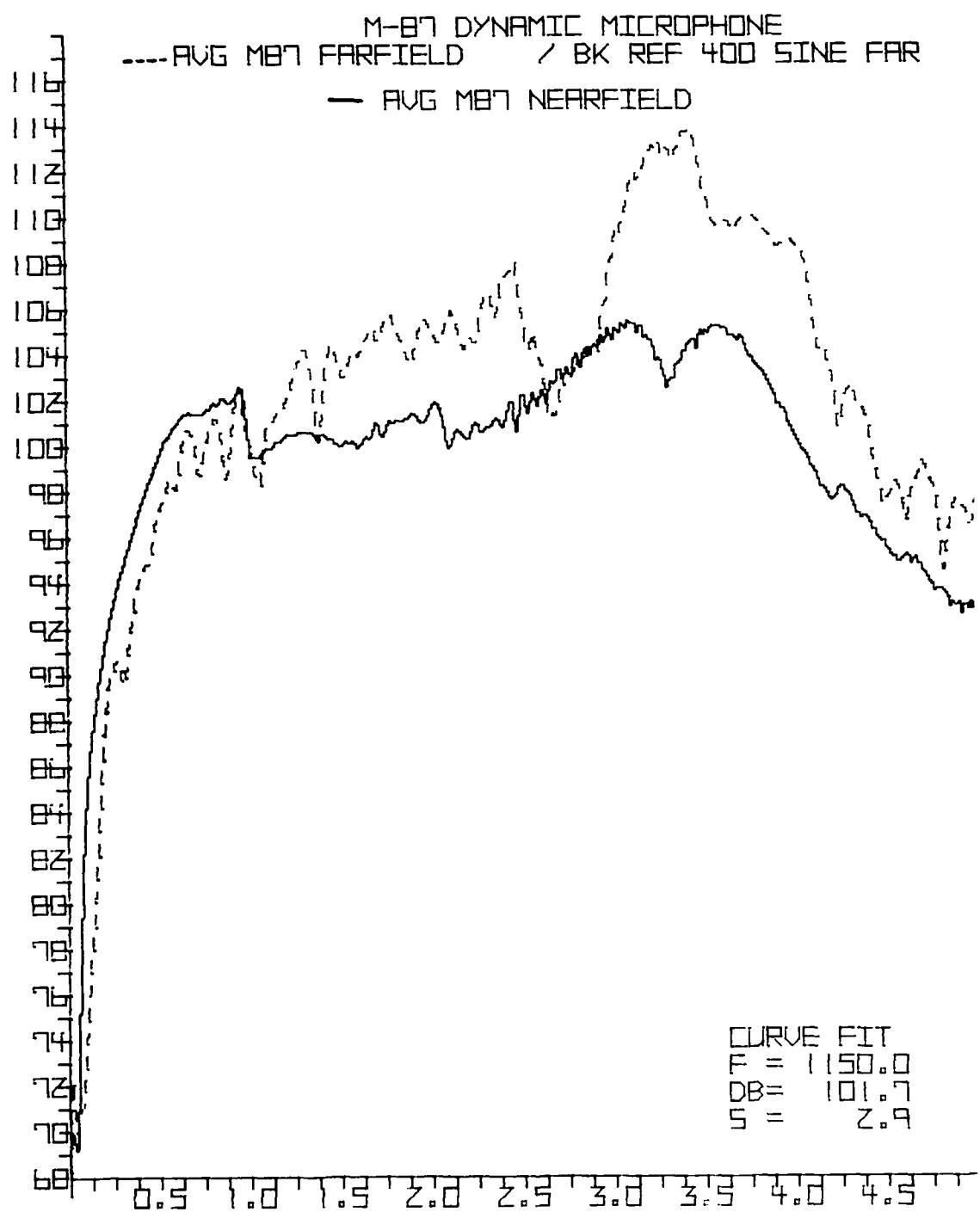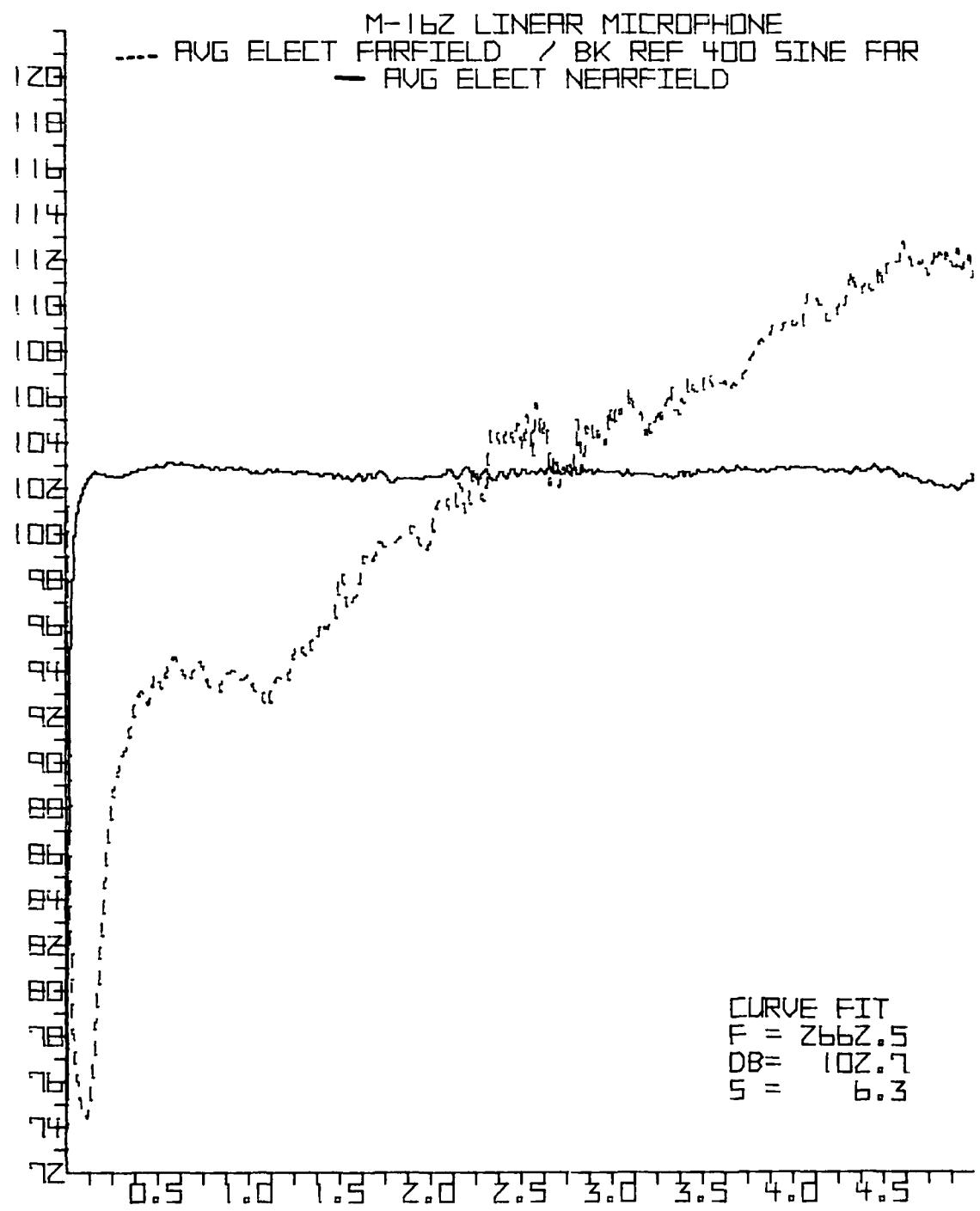
FIGURE 2

FIGURE 3

FIGURE 4

assemblies. These linear earcups provide improvements in noise attenuation over the existing earcups found in the SPH-4 aviator's helmet, as shown in Figure 5. These improvements, while significant, are in addition to improvements in the frequency response on the earphone elements when they are tested inside the new earcups. These improvements can be clearly seen in Figure 6 which compares the frequency response of the old earcup transducer assembly to the earcup transducer assembly of the MK-1564( )/AIC.

5. FUTURE PROGRAMS

a. Voice Recognition and Response for Army Aircraft (VRRAA). The Avionics Research and Development Activity (AVRADA) has just initiated a program entitled "Voice Recognition and Response for Army Aircraft." The VRRAA program will take a phased approach to the introduction of voice I/O equipment into the Army aircraft environment. The first phase of the VRRAA program will utilize the extensive acoustical analysis and simulation facility of AVRADA to systematically evaluate the performance of candidate, off-the-shelf, voice I/O equipments in various Army aircraft noise environments.

Testing in the environmental simulation chamber is expected to begin in February 1981.

The subsequent phases of the VRRAA program will include the interface of selected voice I/O equipments with the AVRADA developed Integrated Avionics Control System (IACS). This will permit the evaluation of the voice I/O equipments in the simulated noise environment while actually performing voice-controlled aircraft radio and frequency selection. As the predictability of the voice I/O equipment in the simulated aircraft noise environment is established, the testing of selected voice I/O equipments will begin in actual aircraft. It is hoped that this testing will provide a baseline of information from which specifications and requirements for the development of Army aircraft unique voice I/O equipment. The aircraft testing will culminate in the integration of voice I/O equipment in AVRADA's System Test Bed for Avionics Research (STAR) aircraft.

The STAR aircraft is a UH-60 Blackhawk helicopter which will be configured to include a 1553 multiplexed data bus and a multiplexed digital audio bus (DMAS). In this environment the voice I/O system will have access to all the audio intercom systems for voice I/O purposes and all the avionics for control applications. Applications testing will be performed in the STAR aircraft to determine which aircraft operational functions would be suitable for voice control and response.

b. Digital Multiplexed Audio System (DMAS). The DMAS program will develop a bus-structured, digitized audio processing and distribution system to achieve maximum reduction in aircraft system wiring effort and cost. DMAS will integrate the communication system operational control functions into the aircraft system standardized bus structure (e.g., MIL-STD-1553( )).

Optical communications links will also be evaluated in DMAS because they can add additional redundancy and further reduce the weight of the aircraft wiring harness over a conventional wire system. Optical systems, however, have their own unique design barriers, which must be overcome to enable their efficient and effective use in the adverse environment of military aircraft.

The DMAS program will investigate various bus structures as well as digitizing techniques to determine the most cost-effective and efficient system for implementation into future aircraft designs and retrofit programs.

DMAS will be designed with a modular structure to permit incorporation of future design improvements with minimum impact on the system.

6. CONCLUSIONS

Outmoded standards and testing techniques must be replaced with modern test procedures which can more accurately test and evaluate the various components of the communications system. These new test procedures must be carefully designed so that they more closely reflect conditions that exist in the actual flight environment.

The total communications system must be investigated in order to improve electronic voice communications in the high noise environment of military aircraft. Improvements made to only one or two elements of the system will most likely have little or no impact on the overall system response characteristics.

7. REFERENCES

a. Giordano, T. and Keane, G., United States Army Electronics Command, The Effect of Helicopter Noise on Communication and Hearing, 1973, ECOM-4140.

b. Mayer, M., United States Army Avionics Laboratory, Acoustic Noise Suppression in Airborne Communications, 1975, ECOM-4336.

c. Lindberg, A. and Mayer, M., United States Army Avionics Laboratory, Communication in the Aircraft (High Noise) Environment, 1976, Proceedings of Aviation Electronics Symposium, Volume 2, Army Aviation Association of America.

d. Lindberg, A. and Mayer M., United States Army Avionics Laboratory, Laboratory Testing Techniques for High-Noise Environment Communications Systems, 1976, AD-A026062, Army Science Conference Proceedings, Volume II, Pages 373-387.

Attenuation Characteristics of MK-1564()/AIC Verses MK-896A/AIC
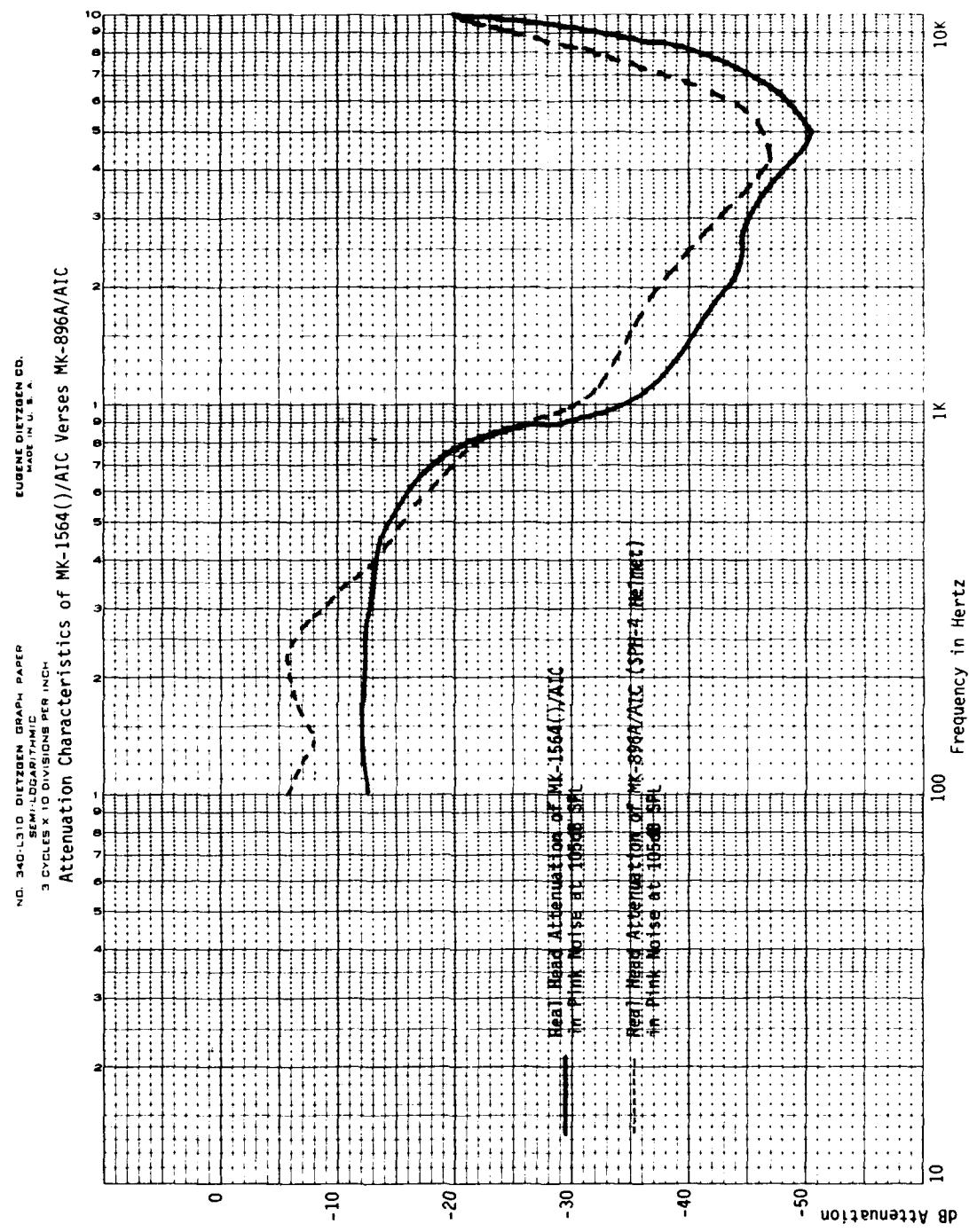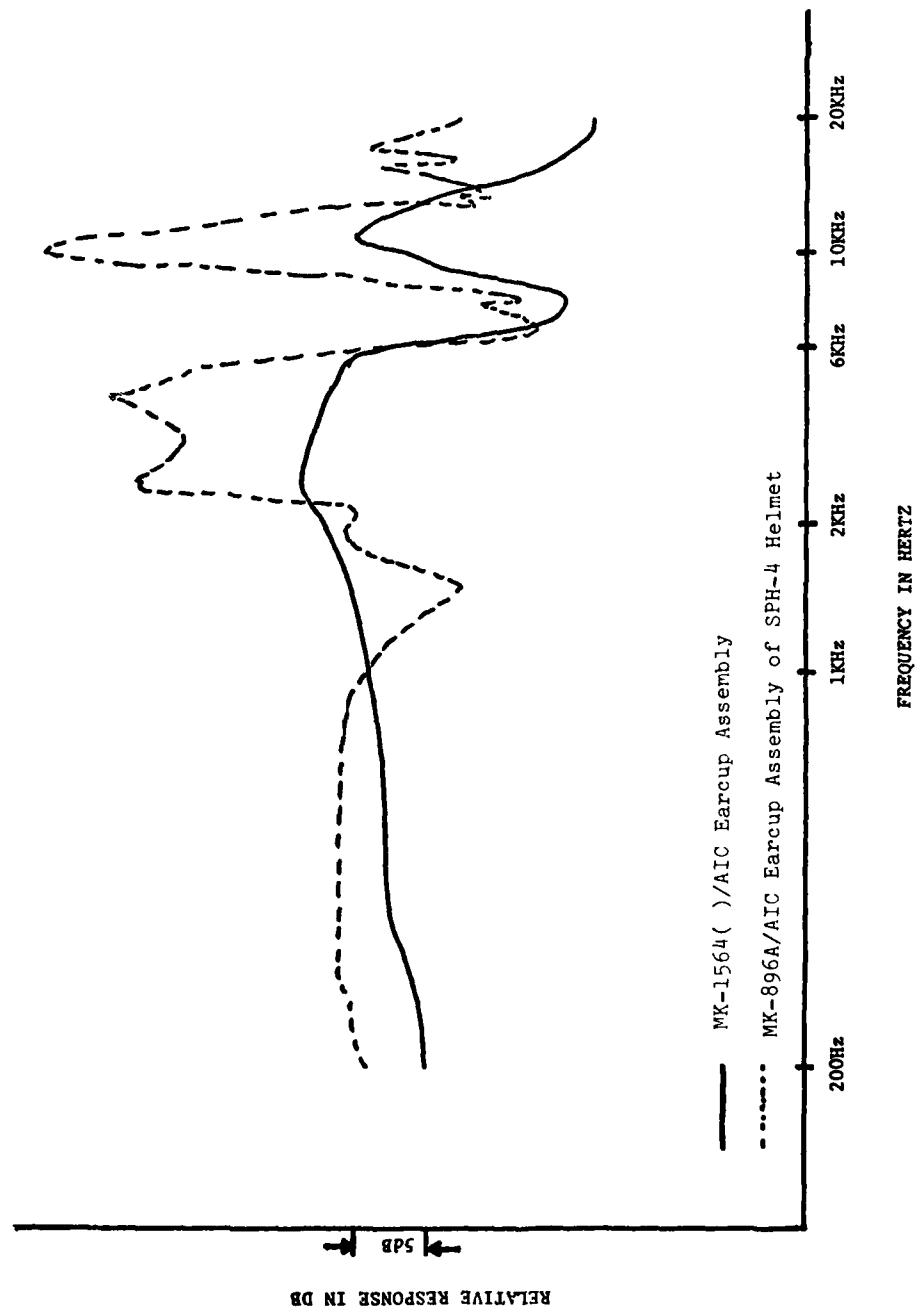
FIGURE 5

FIGURE 6

Comparison of MK-1564( )/AIC Earcup Transducer to the Earcup Transducer of SPH-4 Helmet

e. Mayer, M. and Lindberg, A., United States Army Avionics Research and Development Activity, A Survey of Communications in the High Noise Environment of Army Aircraft, 1978, Conference Proceedings No. 255, Operational Helicopter Aviation Medicine, AGARD/NATO, Pages 46-1 to 46-18.

f. Mayer, M. and Lindberg, A., United States Army Avionics Research and Development Activity, Analysis of Communications in the Noise Environment of Army Aircraft, Jun 79, 97th Meeting of the Acoustical Society of America, Paper No. Q-8.

THE EFFECT OF NOISE ON THE VESTIBULAR SYSTEM.

W.J. Oosterveld, MD.[*], A.R. Polman, MD.[**], and J. Schoonheyt, MD.[**]

[*] University of Amsterdam, Wilhelmina Gasthuis, 1e Helmersstraat 104, 1054 EG Amsterdam, The Netherlands
[**] Medical Department Royal Dutch Airlines, Schiphol, The Netherlands.

Introduction

Acoustic stimulation affects besides the cochlear organ, the vestibular organ too. Tullio[1,2] gave his name to the phenomenon that sound applied to one ear provokes symptoms from the vestibular organ as nystagmus, vertigo and body movements. Tullio assumed that the ampullary cristae had not only a vestibular function, but were involved in the process of directional hearing too. Van Caneghem[3], Dohlman[4,5], Van Eunen, Huizing and Huizinga[6], Huizinga[7,8], Jellinek[9,10] and Hennebert[11] confirmed the findings of Tullio that sound elicited vestibular reactions. Kwee[12] and Tonndorf[13] found this effect in deaf people and Bleeker, Wit and Serenhout[14] in deaf pigeons. Even microphonic activities from the ampullary cristae in pigeons were recorded (Bleeker and De Vries[15,16], Huizinga, De Vries and Vrolijk[17]).

Sound presented at both the ears does not provoke the Tullio-phenomenon (Von Békésy[18]), as the reflexes equalize each other.

However, when normal humans are subjected to noise of more than 120 dB, vertigo and nystagmus are provoked. Vestibular disturbances induced by jet-engine noise are described by Titow[19], Mittelmayer[20] and Dickson and Chadwick[21]. Even very loud music is able to arise vertigo (Van Eunen, Huizing en Huizinga[6]). Parker, Von Gierke and Reschke[22] and Harris[23,24] reported not only vestibular symptoms as nystagmus and dizziness but nausea too. Labourers working in an environment with a high level of low frequency sound were found to have an increased sensitivity of the labyrinth(Temkin[25]).

Habermann[26] reported that tinkers with a noise induced hearing loss suffered in many cases from balance disorders and had a spontaneous nystagmus. These findings were confirmed by Orenbowski[27] in tinkers and by Sirala[28] in shipyard workers. Remarkable is the finding of Bugard[29] of the existence of adrenal gland insufficiency in aircraft factory workers who were subjected to industrial noise for a long time.

Gerlings and De Kleyn[30] described a case in which vestibular irritation, dizziness and nystagmus were elicited by acoustic stimulation.

Inner ear disorders are able to lower the threshold of the vestibular sensory epithelium for acoustic stimulation to such an extend that even normal environmental noise provokes vertigo. This was described by Tullio[2] in a patient with labyrinthitis. After stapedial surgery (Menzio[31]) and in some cases of Ménière's disease (Naito[32], Lange[33], Lindsay[34]) this phenomenon was found. Spitzer and Ritter[35] described it in a patient with a latero-basal fracture of the skull.

Explanations of the working mechanism of the phenomenon are discussed by several authors. Kacker and Hinchcliffe[36] suggest that movements transmitted through the connection of the ossicles with the membraneous labyrinth are the cause. Kwee[37] assumes that the phenomenon originates in a malfunction of the stapedial muscle. Benjamins[38] stressed the fact that an intact middle ear apparatus is necessary. Moulonguet and Poncet[39] assumed that intensive noise can cause sudden contraction of the intra-tympanic muscles, resulting in sudden jerks of the stapes which sets the perilymph of the semicircular canal in motion.

Cochlea and labyrinth are parts of one system with many similarities in their working mechanism. The fluid circulation system of both the organs has strong ties. Phylogenetically is the cochlea a part of the vestibular organ with a high differentiation in its structure.

The aim of the present study was to find out if men with a noise induced hearing loss had detectable functional changes in the functioning of their vestibular system. The study was carried out in a group of 29 technicians of the engine maintenance department of Royal Dutch Airlines. They all had worked in the rather noisy environment of the workshop for more than five years and all had a hearing loss on both ears of more than 40 dB on the tone audiogram at the frequency of 4000 Hz.

Methods

All technicians had a history free from audiovestibular disorders in the period preceding their present job. Three subjects had complaints about irregular existing dizziness spells. For the present study they all were subjected to hearing tests and an extensive vestibular examination.

Hearing tests were conducted by means of tone audiometry and speech audiometry.

The vestibular examination was conducted with the aid of electronystagmography. With this technic eye movements in the horizontal plane were recorded. The examination consisted of the observation of the presence of spontaneous nystagmus and positional nystagmus in any of four different positions, i.e. supine, prone, left lateral and right lateral.

Extensive cervical movements were carried out by the subjects in order to provoke a cervical nystagmus. Furthermore eye movement tests were conducted as the observation of the presence of a fixation nystagmus and a gaze nystagmus. Eyetracking tests consisted of the pendular eyetracking test, where the visual fixation of the subject on an oscillating target is studied, and the optokinetic nystagmus test in which visual stimulations are given with velocities of 30°, 60°, 90° and 120°, in a clockwise- as well as in a counterclockwise direction. The eyetracking tests were conducted with the vision of both eyes together as well as with the vision of each eye separately.

All cases were subjected to a rotation test by means of a torsion swing which provokes an alternating rotational acceleration in the horizontal plane with an oscillation time of 20 seconds.

Results

According to the grade of their hearing loss, at the frequency of 4000 Hz, the subjects were divided into four groups.
In group I the subjects had a hearing loss of less than 50 dB, in group II of 50-60 dB, in group II of 60-80 dB and in group IV of more than 80 dB. The groups consisted respectively of 5, 14, 8 and 2 subjects. The average age of the subjects at the time the study was conducted was 44 years (29-52) in group I, 47 years (22-63) in group II, 47 years (31-62) in group III and in group IV 54 years (53-55 years). (fig.1).

figure 1:

GROUP	HEARING LOSS AT 4000 Hz	NUMBER OF SUBJECTS	AGE Average	AGE Range
I	< 50 dB	5	44	29-52
II	50-60 dB	14	47	22-63
III	60-80 dB	8	47	31-62
IV	> 80 dB	2	54	53-55

The vestibular examinations showed that 18 of the 29 persons (62%) had spontaneous nystagmus with a speed of the slow component exceeding 5°/second. A positional nystagmus exceeding 5°/sec., appearing in three or more positions, was found in 24 subjects (83%). In all cases the nystagmus was direction fixed.
A cervical nystagmus could be provoked in 17 subjects (59%).
The visual fixation test and the eyetracking tests did not show pathology in any of the subjects.
The rotation test aroused in all subjects a nystagmus. In 7 persons (24%) a preponderance of more than 20% for one nystagmus direction appeared.
The caloric nystagmus, the only test which compares the excitability of both the labyrinths, showed a labyrinth preponderance of more than 20% in eight cases. A nystagmus direction preponderance in this test of more than 30% appeared in seven cases. (fig. 2, fig. 3).

figure 2:

Group	Subject	Age	A	B	C	D	E	F	G
I	1	22	+	+	−	−	−	−	+
	2	48	−	+	−	−	+	+	−
	3	41	+	+	+	−	−	−	−
	4	29	−	+	−	−	+	+	−
	5	50	+	+	+	−	+	−	−
II	6	39	+	+	+	+	−	+	−
	7	39	+	+	+	−	−	−	+
	8	51	+	+	+	−	−	+	+
	9	53	+	+	+	−	−	−	+
	10	63	−	+	−	−	−	−	−
	11	32	+	+	−	−	−	−	−
	12	53	+	+	−	−	+	−	−
	13	62	+	+	+	+	+	+	−
	14	34	+	+	+	−	+	+	−
	15	44	−	−	−	+	−	−	+
	16	41	−	+	−	+	−	−	−
	17	56	+	+	−	−	−	−	−
	18	59	−	−	+	+	−	−	−
	19	61	+	+	+	−	−	−	+
III	20	53	+	+	+	−	−	−	+
	21	44	−	+	−	−	−	−	−
	22	31	−	−	+	+	−	−	+
	23	35	+	+	−	−	−	−	−
	24	50	+	+	+	+	+	−	−
	25	62	−	+	+	−	−	+	−
	26	51	+	+	+	+	−	−	+
	27	59	−	+	+	−	+	−	−
IV	28	52	−	−	+	−	+	−	+
	29	55	+	−	−	−	−	−	−

A = spontaneous nystagmus: more than 5°/s = +; B = positional nystagmus: more than 5°/s in 3 positions = +; C = cervical nystagmus: presence = +; D = rotation test: nystagmus preponderance more than 20% = +; E = caloric vestibular test: labyrinth preponderance more than 20% = +; F = caloric vestibular test: nystagmus preponderance more than 30% = +; and G = complaints about vertigo = +.

figure 3:

Discussion

Spontaneous and positional nystagmus as well as the presence of nystagmus suggests pathology in the vestibular system. According to Bos, Oosterveld and Philipszoon[40] appears a spontaneous and positional nystagmus without a pathological value in 5% of the normal population. A difference in excitability of more than 20% without any detectable pathology excists in less than 0,5% of the normal population (Hamersma[41]). The fact that only 2 subjects were free from this symptoms strongly suggests irritative conditions of the balance mechanism in all subjects. The rather high percentage of persons with a cervical nystagmus can be explained by the fact that their type of work forced them many times to extensive head and neck movements, causing slight changes in the cervical structures. In most cases these patients remain free from subjective symptoms (Janzen[42], Von Manz[43]). A cervical nystagmus has nothing to do with noise induced vestibular damage.
The eye movements and tracking tests were conducted normally by all men, which means that there is no central pathology within the oculomotor system.
The rotation test showed pathology in seven subjects, caused evidently by the rather strong spontaneous nystagmus.
The caloric nystagmus test indicated in eight men a difference in excitability between the labyrinths, in

four cases showed the right labyrinth and in the others the left labyrinth a diminished excitability.
An explanation can be the assumption that the labyrinths were not equally damaged.
In the groups I, II and III a high number of pathological findings appear. Only two men were free from
spontaneous and positional nystagmus. The subjects of the groups III and IV showed a higher percentage
of abnormal findings than the subjects of group I, which points to the fact that with an increasing hearing
loss the vestibular damage increases too.
No evidence consists that special parts of the vestibular organs are affected especially. Theoretically the
statolith organs are expected to suffer more than the canal system.
The lack of complaints as dizziness and vertigo in most subjects is most likely to depend on the well
developed property of the vestibular system to adapt to slowly arising damage. This means that much harm
can be done to the vestibular organs without existence of major complaints. The findings lead to the
assumption that a noise induced hearing loss is accompanied by a rather severe damage to the vestibular
system.
The middle ear protects the cochlea from noise damage by means of a reflectory contraction of the middle
ear muscles. This protection concerns only the sound frequencies up to 2000 Hz, and not higher frequencies
(Sterner[44]). Noise up to 85 dB is harmless, but when the loudness is 90 dB and higher a hearing loss will
develop (Yaffe, Jones and Weiss[45]).
After the first exposures to damage noise the cochlea has the capacity to recover from the developing
hearing loss, especially in the lower frequencies (Russi[46]). However, when the noise induced hearing loss
increases this damage is irrevocable.
The findings in this study prove the fact that excessive noise damages besides the cochlea also the labyrinth.
The avoidance of noise exposure by means of ear defenders protects not only the hearing from damage, but
preserves the vestibular function too.

References
1. Tullio, P.: Das Ohr und die Entstehung der Sprache und Schrift.
Munchen, Urban & Schwarzenbeck, 1929.
2. Tullio, P.: Démonstration des méthodes pour la stimulation acoustique des caneaux semicirculaires.
Acta Otolaryngol. 26: 267, 1938.
3. Cameghem, D. van: La surdité subite.
Acta Otorhinolaryngol. Belg. 12: 5, 1958.
4. Dohlman, G.: Discussion of "Der Resonanz von Tönen in den Meermuscheln".
P. Roorda. Acta Otolaryngol. 15: 324, 1931.
5. Dohlman, G., and K. Money: Experiments on the Tullio vestibular fistula reaction.
Acta Otolaryngol. (Stockh.) 56: 271, 1963.
6. Eunen, A.J.H. van, Huizing, H.C. and E. Huizinga: Die Tulliosche Reaktion im Zusammenhang mit der
Funktion des Mittelohres.
Acta Otolaryngol. (Stockh.) 31: 265, 1943.
7. Huizinga, E.: Ueber die Schallreflexe von Tullio.
Pflügers Arch. 234: 665, 1934.
8. Huizinga, E.: Weitere Untersuchungen über die Schallreflexe Tullios.
Pflügers Arch. 237: 494, 1936.
9. Jellinek, A.: Akustische Reflexen an Tauben nach isolierten Verletzung der knöckernen Bogengänge.
Mschr. Ohrenheilk. 62: 241, 1928.
10. Jellinek, A.: Ueber "akustische" Reflexe an labyrinthlosen Tauben.
Mschr. Ohrenheilk. 62: 847, 1928.
11. Hennebert, P.: Nystagmus audiocinétique.
Acta Otolaryngol. 51: 412, 1960.
12. Kwee, H.L.: The occurence of Tullio phenomenon in congenitally deaf children.
J. Laryngol. Otol. 90: 501, 1976.
13. Tonndorf, J.: Eine akustische ausgelöste Vestibularisreaktion.
Klin. Wschr. 26: 149, 1948.
14. Bleeker, J.D., Wit, H.P. and H. Serenhout: Evidence for sound perception with the labyrinth.
Acta Otolaryngol. 89: 76-84, 1980.
15. Bleeker, J.D. and H.L. de Vries: The microphonic activity of the labyrinth of the pigeon.
Part I: The Cochlea.
Acta Otolaryngol. (Stockh.), 37: 287, 1949.
16. Bleeker, J.D. and H.L. de Vries: The microphonic activity of the labyrinth of the pigeon.
Part II: The response of the cristae in the semicircular canals.
Acta Otolaryngol. (Stockh.), 37: 298, 1949.
17. Huizinga, E., de Vries, H.L. and J.M. Vrolijk: Analysis of the microphonic activity of the labyrinth
of the pigeon into the contributions of the various parts.
Acta Otolaryngol. (Stockh.), 39: 372, 1951.
18. Békésy, G. von: Ueber akustische Reizung des Vestibularapparatus.
Pflügers Arch. 236: 59, 1935.
19. Titow, A.I.: Mediz. Wehrmachtsjournal, 6: 341, 1938.
20. Mittelmaier, R.: Gleichgewichtsstörung als Folge einer Lärmbeschädigung durch Düsentriebwerke.
Z. Laryng. Rginol. 49: 640, 1970.
21. Dickson, E.D.D. and D.L. Chadwick: Observations on disturbances of equilibrium and other symptoms
induced by jet engine noise.
J. Laryng. 65: 154, 1951.
22. Parker, D.E., Von Gierke, H.E. and M.F. Reschke: Studies of acoustical stimulation of the vestibular
system.
Aerospace Med. 39: 1321, 1968.
23. Harris, C.S.: Effects of acoustical stimuli on the vestibular system.
Technical Report 71-58, Aerospace Medical Research Laboratory, Wright-Patterson AFB, OH, 1971.
24. Harris, C.S.: Effects of increasing intensity levels of intermittent and continuous 1000 Hz tones on
human equilibrium.
Technical Report 72-11, Aerospace Medical Research Laboratory, Wright-Patterson AFB, OH, 1972.
25. Temkin, J.D.: Die Berufstaubheit.
Monatschr. Ohrenheilk. 67: 705, 1931.

26. *Habermann, J.*: Ueber die Schwerhörigkeit der Kesselschmiede.
    Arch, Ohrenheilk. 30: 1, 1890.
27. *Orenbowski, N.S.*: Die otiatrische Untersuchung des Kesselschmiede in den Hauptwerkstätten der Trans-
    kaukaischen Eisenbahn.
    Acta Otolaryngol. 10: 197, 1927.
28. *Sirala, U. and E. Lahikainen*: Studies of deafness in shipyard labourers.
    Acta Suppl. 67: 107, 1948.
29. *Bugard, P.*: Presse méd. Paris, 63: 24, 1955.
30. *Gerlings, P.G. and A. de Kleyn*: Vestibular nystagmus caused by acoustic stimulation.
    Acta Otolaryngol. (Stockh.), 30: 150-155, 1942.
31. *Menzio, P.*: I riflessi di Tullio in sogeti operati di fenestrazione labirintica,
    Otol. Ecc. Ital. 20: 168, 1952.
32. *Naito, T,*: Three cases of Méniére's disease showing Tullio's reaction.
    Otol. Fukuoka I, 249 zit. nach Zbl. Hals.-, Nas.- u. Ohrenheilk. 54: 265, 1955/56.
33. *Lange, G.*: Das Tullio-Phänomen und eine Möglichkeit seiner Behandlung.
    Arch. Ohr.-, Nas.- und Kehlk.-Heilkunde 187: 643, 1966.
34. *Lindsay, H.R.*: Healing of experimental labyrinthine fistulae.
    Arch. Otolaryngol. 46: 584, 1947.
35. *Spitzer, H. and K. Ritter*: Ein Beitrag zum Tullio Phänomen.
    Laryngol. Rhinol. 58: 934-936, 1979.
36. *Kacker, S.K. and R. Hinchcliffe*: Unusual Tullio phenomena.
    J. Laryngol. Otol. 84: 155-166, 1970.
37. *Kwee, H.L.*: A case of Tullio-phenomenon with congenital middle-ear abnormalities.
    ORL (Basel), 34: 145, 1972.
38. *Benjamins, C.E.*: Les réactions acoustiques de Tullio chez l'homme.
    Acta Otolaryngol. (Stockh.) 26: 91-249, 1938.
39. *Moulonguet, M. and J. Poncet*: Un cas de vertige provoqué par le bruit.
    Ann. Otolaryngol. (St. Louis) 64: 84, 1947.
40. *Bos, J.H., Oosterveld, W.J., Philipszoon, A.J., Vozza, J. and S. Zelig*: On pathological spontaneous
    and positional nystagmus.
    Pract. Oto-rhino-laryngol. 25: 282, 1963.
41. *Hamersma, H,*: The caloric test.
    Thesism Amsterdam, 1957.
42. *Janzen, R.*: Neurologisch-klinisches Bild des cervicalen Vertebralsyndroms.
    In: F. Reischauer, De cervicalen Vertrbralsyndrome. Thieme, Stuttgart, 1955.
43. *Manz, A. von*: Zur Pathologie und Begutachtung des Halswirbelsäulensyndroms.
    H.N.O., 6: 371, 1956/57.
44. *Sterner, J.H.*: Are we ready for standards in noise?
    Industr. Med. Surg. 20: 99, 1951.
45. *Yaffe, C.D., Jones, H.H. and E.S. Weiss*: Industrial noise and hearing loss in a controlled population.
    First report of findings.
    Aiha J. 19: 296, 1958.
46. *Russi, U.*: Ueber Lärmschädigungen in industriellen Betrieben und ihre Verhütung.
    Pract. Otorhinolaryngol. 8: 71, 1946.

Summary
Many authors have reported about the stimulation of the vestibular organs by means of noise.
In this study the vestibular function was examined in 29 technicians, all with an industrial type of
hearing loss. Eighteen showed a spontaneous nystagmus and 24 had a positional nystagmus exceeding a
velocity of 5°/second in three or more positions. A cervical nystagmus could be provoked in 17 subjects.
In 7 persons, the rotational test proved a nystagmus preponderance of more than 20%. A difference in
excitability between the labyrinths of more than 20% appeared in 7 cases. The tests for central vestibular
disorders did not show pathology in any of the subjects. According to their hearing loss, the subjects
were divided into four groups. There was no correlation between the grade of hearing loss and the
vestibular function disturbance. An explanation can be sought in the adaptation capability of the
vestibular system. All technicians showed pathology in one or more of the vestibular tests. This means
that noise exposure not only damages the cochlea, but threatens the vestibular organs too.

DISCUSSION

DR. J.D. MOSKO (UNITED STATES)

Did you have any means of determining whether the onset of the vestibular damage lagged the onset of the hearing loss or whether it paralleled the onset of the hearing loss?

From which group did the two subjects who did not exhibit symptoms come?

DR. W.J. OOSTERVELD (NETHERLANDS)

Because the subjects were investigated only one time, and as they had all a hearing loss of more than 40 dB the onset of the vestibular damage could not be detected.

Both subjects belonged to group I.

MAJ. G. DE HEYN (BELGIUM)

1. What was the age group of the subjects you studied?
2. Were the patients you selected at the beginning complaining of vertigo and whether you selected them on that basis?
3. Did you carry out the same study with subjects that had no heavy loss in similar age groups.

DR. W.J. OOSTERVELD (NETHERLANDS)

1. The age group ranged from 23 years up to 58 years.
2. We performed the same study on a control group composed of workers of the same age.
3. No. We did not have a control group but we know from other studies the numbers of people with a hearing loss in these environments so with people without hearing loss, we know exactly in what amount the ear responded with spontaneous nystagmus, positional nystagmus, surface nystagmus, and what is normal in a population to have a diminished excitability of the labyrinth, but in this case it was not figured out specifically.

AIR COMMODORE P.F. KING (UK)

Please comment on the significance of the findings in terms of the assessment of those men who are claiming compensation for noise-induced deafness.

DR. W.J. OOSTERVELD (NETHERLANDS)

That is a very difficult question but I told you that the vestibular damage is arising very slowly and it was that we were carrying out this study and then we found out that many of these people suffered from vestibular damage, but the vestibular organ is a sense organ with a very high grade of adaptation possibilities. So this means that if there is very heavy damage, that in many cases the patients do not have any complaints. You have seen our study that there are only 10 cases of the 29 subjects that there were some complaints of dizziness or vertigo but all the other signs and symptoms were found by the study and were not based on any complaints of these persons.

MR. R.T. CAMP (UNITED STATES)

I first observed alot of nystagmus among deaf subjects in an experiment that Graybiel conducted, and in which I assisted at Gauladet School for the Deaf in Washington. Perhaps some of you are familiar with the data. We subjected alot of the deaf subjects to high sound levels. We created a 160 dB in a tube and of course, under these conditions, you could assume that you have a sonic colaric test because hands could warm up at the highest levels and it really got me thinking about the theory of what is happening. Is it low-frequency turbulence, the low-frequency filtering into the vestibular system? We must have a physical theory of the physiological and physical correlates to fit a low-frequency stimulation. As a result, some years later, I did some exploratory experiments with pigeons and found there was a suggestion of resonance at 2 Hz with a unilateral input with a 2 Hz pressure field.

Are you implying that the high frequency can filter through also without being transformed in some way or other?

DR. W.J. OOSTERVELD (NETHERLANDS)

The cause is vibration. Up to 2000 Hz the vibration is levelled out by the action of the middle ear muscles. Above 2000 Hz there is no protection of the hearing organ or the vestibular organ. This causes the damage.

THE SPEED OF RESPONSE TO SYNTHESIZED VOICE MESSAGES

by
Dr John L. Wheale
Senior Psychologist
Flight Skills Section
RAF Institute of Aviation Medicine
Farnborough, Hampshire

SUMMARY

The experiment evaluates the effectiveness of synthesized cockpit voice warning messages using measures of reaction time (RT). Research has shown that voice messages are comparable to audio warnings and that synthesized voice messages are easily recognisable at low signal to noise ratios. This study evaluates synthesized voice messages by observing their performance in combination with auditory and visual indicators. Four different warning arrangements were used of which three had a Votrax voice component. The four warning systems represented future possible warning combinations for transport aircraft. Subjects had to deal with simulated emergencies whilst performing a psychomotor tracking task and monitoring ATC messages. Thirty commercial pilots took part in the study. Overall the four warning systems were equally effective in terms of RT. However, voice messages had significantly slower RTs than audio warnings. Voice messages and illuminated legends caused significantly less disruption of ATC monitoring than audio warnings. Pilots also consistently cross checked voice and audio inputs with visual indicators. It is suggested that the proliferation of cockpit voice inputs should be avoided until human factor evaluations of simpler but equally effective warning systems has been completed.

INTRODUCTION

Warning indicators should be able to catch attention and inform the recipient of the nature of the emergency. In achieving both of these functions warning indicators should allow for a rapid and easy transition from warning to corrective action. Because of the variety and extent of visual stimuli in the jet transport cockpit indications of aircraft emergencies have often been presented in the auditory channel. Thus, there is a tradition of using warning sounds (audio warnings), like the firebell, which have been shown to be effective warning indicators. With the advent of wide-body transports the number of audio warnings used on the flight-deck has increased considerably. The use of multiple audio warnings continues despite the fact that recent work has shown that a multiplicity of audio warnings can have undesirable side-effects. Often audio warnings are presented at unnecessarily high intensity levels and this can lead to the masking of speech and disruption of thought processes. A recent report has shown that from the human learning viewpoint the optimum number of audio warnings would be four (Patterson and Milroy, 1980).

If the number of audio warnings used on commercial flight decks were to be limited then it seems quite possible that voice warning messages will be used to supplement them. Using a voice message to signal emergencies is not a new idea. However, the use of voice messages is gaining favour as the electronic storage and processing of voice data develops (Simpson and Williams, 1980). Potentially voice messages have all the advantages that are associated with audio warnings. But, voice messages have the added advantage of being directly comprehensible, it is not necessary to learn the meaning of a phrase as it is necessary to learn the meaning of an interrupted horn, for example. Given that voice messages are potentially very useful there remains an issue that concerns the role such messages might take in a centralised warning system. One way of assessing the role of voice messages is to survey previous human factor evaluations of them.

Most research in this area has concentrated on comparing voice messages with audio warning and/or with illuminated legends on a central warning panel (CWP). In all comparisons, voice warnings were at least as good as illuminated legends and in some produced faster reaction times (RT). Lilleboe (1963) evaluated a voice warning system as a supplement to a visual warning display in the VA-3B aircraft. RTs to warnings presented via captions on the visual display were significantly greater than RTs to similar warnings presented via a voice and visual indicator. A study by Reinecke (1971), conducted in a UH-10 helicopter during low-level and cruising flight, also showed that voice messages, when supplementing visual indications of emergencies, produced faster RTs then when visual indicators were used alone.

Particularly revealing were the experiments of Kemmerling et al (1969) which showed that voice messages become advantageous during periods of high task loading. As visual task loading was increased, pilots responded more quickly to voice messages than they did to audio warnings. This RT advantage occurred because the pilots did not scan the annunciator panel when dealing with a voice message but did scan when dealing with an audio warning. However, in a study reported by Bate and Bates (1967), where the primary task was a visually-complex target location problem, no significant differences were found between visual, visual-audio, and visual-voice warnings. Whether pilots would scan the CWP after receiving a voice warning message would not only be a function of task loading but would also depend upon other factors such as training, normal procedures and attitudes toward the reliability of the warning system. In a survey of British commercial airline pilots (Wheale, 1980) 60% said that they would always cross check a voice warning message with its corresponding visual indicator.

Many recommendations for the use of cockpit voice messages were derived from the research outlined above (Birkhan and Steininger, 1971; Brown et al, 1968). However, there are a number of human factor aspects that have not been examined. Consider the format of the voice message and the type of voice that should be used in the cockpit. Much of the early research in this area involved the use of female voices recorded onto magnetic tape. The female voice was presumed to be distinctive because it was a voice quality very rare in airways communication but nowadays this is not so great. Technological advances have also allowed voice messages to be electronically synthesised and these can be used instead of recorded real speech. Simpson and her colleagues (Simpson, 1976; Simpson and Hart, 1977; Simpson and Williams, 1978)

have conducted a series of studies investigating the use of synthesised voice inputs as warning messages. These studies have shown that voice messages produced by a Votrax synthesiser are reliably intelligible, they are acceptable to line pilots, and they produce faster RTs when presented in the form of a sentence. Voice messages presented in a sentence format required less mental processing capacity to respond to than did similar voice messages presented in a key-word format.

Because voice messages would be part of a centralised warning system it seems obvious that the next step in their evaluation should be an investigation of their performance within a warning system where the contribution of voice messages may be varied. The purpose of the experiment to be reported here was to perform such an evaluation using a part-task simulation of flying activity whilst measuring RT to different types of warning indicators. The general intention was to compare a model of a current warning system, consisting of audio warnings and visual indicators, with similar systems that had varying degrees of voice input. The experiment was aimed at testing the hypothesis that the performance of voice messages would vary according to the role they performed within the warning system as a whole. This work was part of a research programme, funded by the Civil Aviation Authority, which examined the role of voice warning messages on the flight decks of commercial transport aircraft.

METHOD OF EVALUATION

Warning Systems

In the experimental study there was for each of the four conditions 24 possible alerts which were presented using visual, audio or voice indicators. The system used was a derivation of the HS 146 warning panel with master caution lights set in the main field of vision and an annunciator panel, which provided detailed information, positioned to the right of a tracking display. The annunciator panel was sub-divided into three groups of Red (6), Amber (14) and White (4) illuminated legends providing emergency, advisory and information alerts respectively. Four warning system designs were used. The basic premise behind the four different warning systems was that each alert would always consist of an illuminated legend and a flashing attention-getting light of the appropriate category centrally located but that these would be selectively supplemented with a voice message or an audio warning (see Table 1). If an alert had an audio

TABLE 1

ARRANGEMENT OF AUDIO WARNINGS AND VOICE MESSAGES
IN THE FOUR EXPERIMENTAL CONDITIONS

Condition	A	Red Alerts (6) supplemented with audio warnings
Condition	B	Red Alerts (6) supplemented with voice messages
Condition	C	Red Alerts (6) supplemented with audio warnings Amber Alerts for Flight Deviations (4) supplemented with voice messages
Condition	D	Red Alerts (6) supplemented with audio warnings Amber Alerts (14) supplemented with voice messages White Alerts (4) supplemented with voice messages

or voice component the onset of the sound was coincident with the onset of the visual indicators. For each alert there was a response button which eliminated the visual indicator but it did not cancel the audio warning or the voice message. The audio warning repeated for five seconds and the voice warning repeated once to give a total duration of five seconds.

The audio warnings used in the experiment were selected from a variety of aircraft and were chosen because of their distinctiveness and generality (see Table 2). The voice messages were produced by a

TABLE 2

EXPERIMENTAL AUDIO WARNINGS: NAME, AIRCRAFT OF ORIGIN AND DESCRIPTION

Warning	Aircraft	Description
Fire	BAC 1-11	Ringing bell
Undercarriage	L 1011	Horn
Take-off Configuration	BAC 1-11	Intermittent horn
Disconnected A/P	B-747	Siren being started repeatedly
Overspeed	BAC 1-11	Clacker

Votrax ML-1 Speech Synthesiser. The wording of the message is outlined in Table 3; the voice messages were in the key-word format as recommended by US line pilots surveyed by Williams and Simpson (1976). Each voice message began with the word "Warning" which was included to alert and to gain attention. The 'white' or 'information' alerts were in a sentence format.

TABLE 3

EXPERIMENTAL VOICE WARNING VOCABULARY

Emergency	Voice Message
Engine Fire	Warning : Fire No. 1 Engine
	Warning:: Fire No. 2 Engine
Stall	Warning : Stall
U/C Unsafe	Warning : Undercarriage Unsafe
Configuration	Warning : Configuration
Autopilot Disconnect	Warning : Autopilot Disconnect
Fuel Pressure Low	Warning : Fuel Pressure Low
Oil Pressure Low	Warning : Oil Pressure Low
Engine Vibration	Warning : Vibration No. 1 Engine
	Warning : Vibration No. 2 Engine
Electrical Smoke	Warning : Smoke Electrical
Fuel Temp High	Warning : Fuel Temperature High
Ice Detected	Warning : Ice Detected
Hydraulic Pressure	Warning : Check Hydraulic Pressure
Air Conditioning	Warning : Air Conditioning
Engine Overspeed	Warning : Engine Overspeed
Glide Slope	Warning : Glide Slope
Sink Rate	Warning : Sink Rate
Excessive Air Speed	Warning : Air Speed High
Altitude Alert	Warning : Altitude
Radio Fan Off	The Radio Fan is Off
Emergency Lights	The Emergency Lights are On
Cross Feed Open	The Cross Feed is Open
Flt Recorder Off	The Flight Recorder is Off

TEST PROCEDURE

First, the subjects were briefed about the nature and purpose of the experiment. Then the subjects were introduced to one of the four warning systems, to which they were randomly allocated. Finally, the subjects were familiarised with individual elements of the total experimental task. Each subject listened to all 24 voice messages three times. If the warning system included audio warnings (i.e., conditions A, C and D) the subject studied a learning tape of the warnings until he could identify them all without error on three successive presentations. Additionally, the subjects were exposed to the operation of all 24 alerts on the annunciator panel and were required to check the appropriate response button for each alert. This was followed by a ten-minute practice period on the psychomotor tracking task. The tracking task displayed on a CRT involved keeping the intersection of horizontal and vertical cursor lines in the middle of a stationary central square. Pseudo random input nulled by inputs from a two-axis joystick with suitable linkages enabled the lines to be positioned simultaneously but independently. The program for the task was run on a PDP-8 using a Dec 388 display unit. The screen diameter was 406 mm with an active area equivalent to the length of the cursor lines, i.e., 228 mm. The stationary target square had sides of 30 mm. In preliminary trials four pilots were employed to set the tracking task at an activity level that required constant but not intense concentration and effort.

The experimental task was maintained in an environment of simulated cockpit noise (recorded at the Captain's ear during straight-and-level flight in a B-727) and recorded ATC - Aircraft communications, which had been edited to produce continuous conversation. Inserted into the ATC tape were messages prefixed by a unique call-sign which the subjects were required to acknowledge and respond to. These auditory inputs produced a background noise level of 71 db (linear), which is representative of noise level in current transport aircraft. The voice messages and audio warnings were presented over loudspeakers at 80 and 85 db (linear) respectively.

The experiment lasted for 80 minutes during which time 30 alerts were presented, six alerts being presented twice to maintain stimulus uncertainty. The order of the alerts was randomised. The separation time between alerts averaged 160 seconds, varying randomly between 60 and 260 seconds. The response to an alert was quite simple requiring only the cancellation of the appropriate response button. The response buttons were located on a panel to the left of the subject, though they did not have the same spatial configuration as on the CWP so that subjects could not respond on the basis of a transfer of positional sense from the CWP.

SUBJECTS

Thirty pilots took part in the experiment; seven completed conditions A and C and eight completed conditions B and D. The pilots were current on a wide variety of aircraft types and represented five airlines. The group consisted of 11 Captains and 19 First Officers. The average age was 36 years (range 23-54 years) and the average total flying hours was 6,213 (range 1,300-18,000 hours). All the pilots volunteered to take part in their own time; six pilots were drawn directly from British Airways with the remainder volunteering as a result of the British Air Line Pilots Association's co-operation in the study.

RESULTS

The reaction times to individual alerts in the three categories (Red, Amber and White) across the four conditions are shown in Figure 1. The reaction times showed a great deal of variability and some alerts produced reaction times that differed considerably from the group mean. To simplify analysis the RT data were averaged across warning categories and experimental conditions (see Figure 2). Also, prior to analysis of variance, all the RT data were subjected to a log-transformation to normalise the distribution. The mean reaction time for alerts in the Amber and White categories was not significantly different (F 3,226 = 0.5934; p>.1) and so data for these categories were pooled. Figure 2, suggests that condition A, where audio warnings were used for Red alerts and illuminated legends on the CWP were used for the Amber and White categories, produced the lowest average reaction time. However, there was no significant difference in RT according to the mixture of warning types across the four conditions.

With Red alerts, voice messages produced a slower reaction time than audio warnings (F 3,26 = 3.31; p<.05). As a category the Red alerts were not uniform in so far as the emergency, 'CONFIGURATION', produced a RT which differed significantly (F 5,130 = 3.05; p<.05) from the other Red alerts. Additionally, for Red alerts there was a significant interaction between emergencies and indicator type (F 15,130 = 2.13; p<.05). This interaction was present because the RT for the Fire warnings was similar for both voice messages and audio warnings whereas there was a divergence of RTs for the other Red alerts (see Figure 3).

When Amber and White alerts were presented the RTs were statistically identical whether or not they were presented by voice or via the CWP (see Figure 4). There was a significant difference in RT according to which alert was being presented (F 17,440 = 9.19; p<.01) but this effect did not interact with the type of alert used. The alerts that behaved as if they were a separate group were AIR CONDITIONING, OIL PRESSURE LOW, ICE DETECTED, and ENGINE OVERSPEED.

During the experiment subjects had to respond to messages prefixed by a set of call-signs. Although the messages were randomly distributed in relation to the various alerts, occasionally the two inputs coincided. Table 4 shows the number of call-sign messages received in the period between the onset and cancellation of an alert.

TABLE 4

WARNING CATEGORY AND RESPONSE TYPE FOR CALL-SIGN
MESSAGES RECEIVED BETWEEN ALERT ONSET AND RESPONSE

Warning Category	Message Acknowledged	Message Not Acknowledged
CWP	23	2
Voice Message	10	1
Audio Warning	4	5

This table shows that audio warnings were different from voice messages or alerts on a CWP in so far as they were significantly more likely to interfere with simultaneous ATC activity (Chi Square = 10.99; p<.01). The ability of audio warnings to interfere with RT does not persist beyond the cancellation of the alert as is shown in Table 5 where the frequency of message acknowledgement during 15 seconds following the onset of an alert is summarized. The difference between the three warning types did not reach

TABLE 5

WARNING CATEGORY AND RESPONSE TO CALL-SIGN
MESSAGES IN THE 15 SECONDS AFTER ALERT ONSET

Warning Category	Message Acknowledged	Message Not Acknowledged
CWP	32	4
Voice Message	13	4
Audio Warning	7	5

significance (Chi Square = 3.8; p>.05). However, the probability of ATC being disrupted was slightly higher than normal; of the 625 call-sign messages presented that did not coincide with the presentation of

an alert only one message was not acknowledged.

On the basis of data reported by Hick (1952) it was to be expected that RT would increase with the number of warnings in a category (i.e., set size), but no such relationship was found in the present experiment. However, the data suggest that RT may possibly be a function of urgency and set size. Analysis by rank order of the RTs for the individual alerts showed that Red items (high urgen.., set size of 6) produced the fastest RTs with White items (low urgency; set size of 4) producing slower RTs and Amber items (medium urgency; set size of 14) the slowest RTs.

DISCUSSION

The results show that reaction time data did not differentiate in a statistically significant manner between the four warning systems used in the study. It would thus appear that a warning system consisting of audio warnings for Red alerts and illuminated legends for Amber and White alerts is essentially no different to a warning system in which Amber and White alerts are additionally indicated by synthesised voice messages. Indeed, of the 24 different alerts no particular combination of audio, voice and visual indication shows any superiority in respect to reaction times.

However, reaction time cannot be the sole criteria by which a choice amongst warning systems is made. For example, auditory inputs have been considered as good warning indicators because their onset can focus rapidly the listeners attention. In addition, another reason to use auditory warnings is because they allow the pilot to go straight from alert to corrective action minimising the visual workload of cross-checking in a CWP. Kemmerling et al (1969) noted this specific advantage of voice messages, but, in the present study 93% of the pilots cross referred to the CWP before responding when they received either an audio warning or a voice message. In a subsequent interview 57% of the pilots said that they would always cross-check any auditory input with the CWP before responding. This strategy of immediately cross-checking any sound with the CWP is consistent with the use of audio warnings or voice messages merely as attention getters. Indeed, several pilots did comment that this was, and always would be, their strategy for dealing with alerts.

Within the group of Red or 'immediate-action' alerts the data indicate that audio warnings produce faster response times than do voice messages (see Figure 3). This difference is shown clearly in the use of the CONFIGURATION alert where the response to the voice message was slower by over a second. Figure 3 indicates that the CONFIGURATION alert may be exceptional and analysis shows that the RTs to this alert were in fact significantly different from other Red alerts (p<.01). This difference may be partly due to the clarity of the voice message for the CONFIGURATION alert. A high proportion of pilots observed that the pronunciation of the message was not normal. The elevated reaction time for the CONFIGURATION alert was possibly due to the perceptual conflict generated by a clear and intelligible message which had an unusual rhythm and intonation. This problem is a potential source of difficulty for computer synthesised voice messages, which might be overcome by the use of real voices in digital form. However, the problem highlighted by the CONFIGURATION alert indicates that in general the use of voice warning messages may create difficulties as would be the case when a warning message is delivered to a multinational crew, for example.

Another interesting effect was the way in which the RT varied according to the urgency level of the alert; for Red alerts, voice messages produced an RT two seconds faster than similar voice inputs did for Amber or White alerts (see Figure 2). Especially noteworthy is the fact that the voice messages using the sentence format did not produce significantly longer RTs than those produced by shorter voice messages using the key-word format. Reaction time may vary with urgency level because of the operation of two independent factors. First, the urgency level of the alerts used in this study was intermixed with group size; there were six alerts in the Red category and fourteen alerts in the Amber category. Hick (1952) showed that RT to a stimulus will vary according to the number of possible alternatives in the set, thus, choice reaction time will increase with set size. Second, the pilots in this study will almost certainly have transferred to the experiment from their normal experience a differential response set which was based on their flight experience such that they reacted more quickly to the more urgent warning. In this respect it might have been advantageous to use naive subjects. The transfer of a specific response set cannot account for the tendency to react slowly to the AIR CONDITIONING, OIL PRESSURE, ICE DETECTED and ENGINE OVERSPEED alerts. Post hoc analysis revealed that these alerts were grouped in the centre of the response panel along with other alerts that tended to produce slower RTs. It would appear that the uniqueness of the above alerts is due to an artefact of the system design and not due to the nature of the emergencies they represent.

The criterion of reaction time may be necessary but not sufficient to select between warning systems. Several observations on this study confirmed that alerts have a disruptive effect on ATC communications. Indeed, a systematic record of these occurrences showed that audio warnings produced the greatest statistically significant disruption. However, voice messages were clearly less disrupting than audio warnings, and the former caused no more disruption than did visual indicators when presented by themselves. These results are consistent with the many comments pilots have made about audio warnings and their undesirable side-effects.

This experiment was carried out with the intention of using the results to design an 'ideal' warning system with audio warnings and voice messages to supplement the visual indicators on a CWP. The results have shown that reaction time may be useful in evaluating warning systems but it cannot be the exclusive means of differentiating between proposed alternatives. In fact, many factors will interact to influence the effectiveness of any warning system. These include the effect of the warning system on concurrent activities, the influence of the number of warnings (faults, malfunctions) in any warning group, and how the warnings will be used on the flight deck, whether voice messages will merely be used as attention-getters. Because of the variety of factors that can influence the effectiveness of a warning system it is perhaps tempting to advocate the design of a quite simple central warning system and only increase its complexity with voice messages and/or audio warnings when these offer a demonstrable advantage. However, such a design philosophy may be too cautious because of the vital role centralised warning systems have to play in modern aircraft.

ACKNOWLEDGEMENT

REFERENCES

BATE, A.J. (1969). Cockpit Warning Systems Comparative Study. Technical Report AMRL-TR-68-193, May. Wright Patterson AFB, Ohio; Aerospace Medical Research Lab, Air Force Systems Command.

BIRKHAN, G. and STEININGER, K. (1971). Psychological Evaluation of Voice Warning Systems in Airplanes. DFVLR, Institut für Flegmedizin. Abt Luftfahrtpsychologie, Hamburg.

BROWN, J.E., BERTONE, C.M. and OBERMAYER, R.W. (1968). Army aircraft voice-warning system study. US Army Technical Memorandum 6-68.

HICK, W.E. (1952). On the Rate of Gain of Information. Quarterly Journal of Experimental Psychology, 4, p 11-26.

KEMMERLING, P., GEISELHART, R., THORBURN, D.E. and CRONBURG, J.G. (1969). A comparison of voice and tone warning systems as a function of task loading. USAF Technical Report ASD-TR-69-104.

LILLEBOE, M.L. (1963). Final Report. Evaluation of Astropower Inc. Auditory Information Display Installed in the VA-3B Airplane. Naval Air Test Centre Report ST 31-22R-63, AD8031823.

PATTERSON, R. and MILROY, R. (1980). Auditory warnings on civil aircraft: the learning and retention of warnings. MRC APU, Cambridge, England. (To be published).

REINECKE, M. (1971). Experiences with a Vocal Warning System in the UD-1D Helicopter. Paper presented to 6th Meeting of the German Study Group for Flight Data Systems (DSF), Hamburg.

SIMPSON, C.A. (1976). Effects of linguistic redundancy on pilots comprehension of synthesised speech. Twelfth Annual Conference on Manual Control NASA TM x-73, 170, 294-308.

SIMPSON, C.A. and HART, S.C. (1977). Required attention for synthesised speech perception for two levels of linguistic redundancy. Journal of the Acoustical Society of America, 61, Supplement 1, S7/D3.

SIMPSON, C.A. and WILLIAMS, D.H. (1978). The effects of an alerting tone and of semantic context on pilot response time for synthesised speech warnings in a simulated air transport cockpit. MCI Report No 78-001. NASA Ames, California.

SIMPSON, C.A. and WILLIAMS, D.H. (1980). Response Time Effects for Alerting Tone and Semantic Context for Synthesised Voice Cockpit Warnings. Human Factors, 22 (3), p 319-330.

WHEALE, J.L. (1980). Pilot Opinion on Flight Deck Voice Warning Systems. In: Human Factors in Transport Research, Vol. 1, p 88-96, Ed. D.J. Oborne and J.A. Levis. Academic Press: London.

WILLIAMS, D.H. and SIMPSON, C.A. (1976). A systematic approach to advanced cockpit warning systems for air transport operations: line pilot preference. NASA Aircraft Safety and Operating Problems Conference. NASA SP 416.
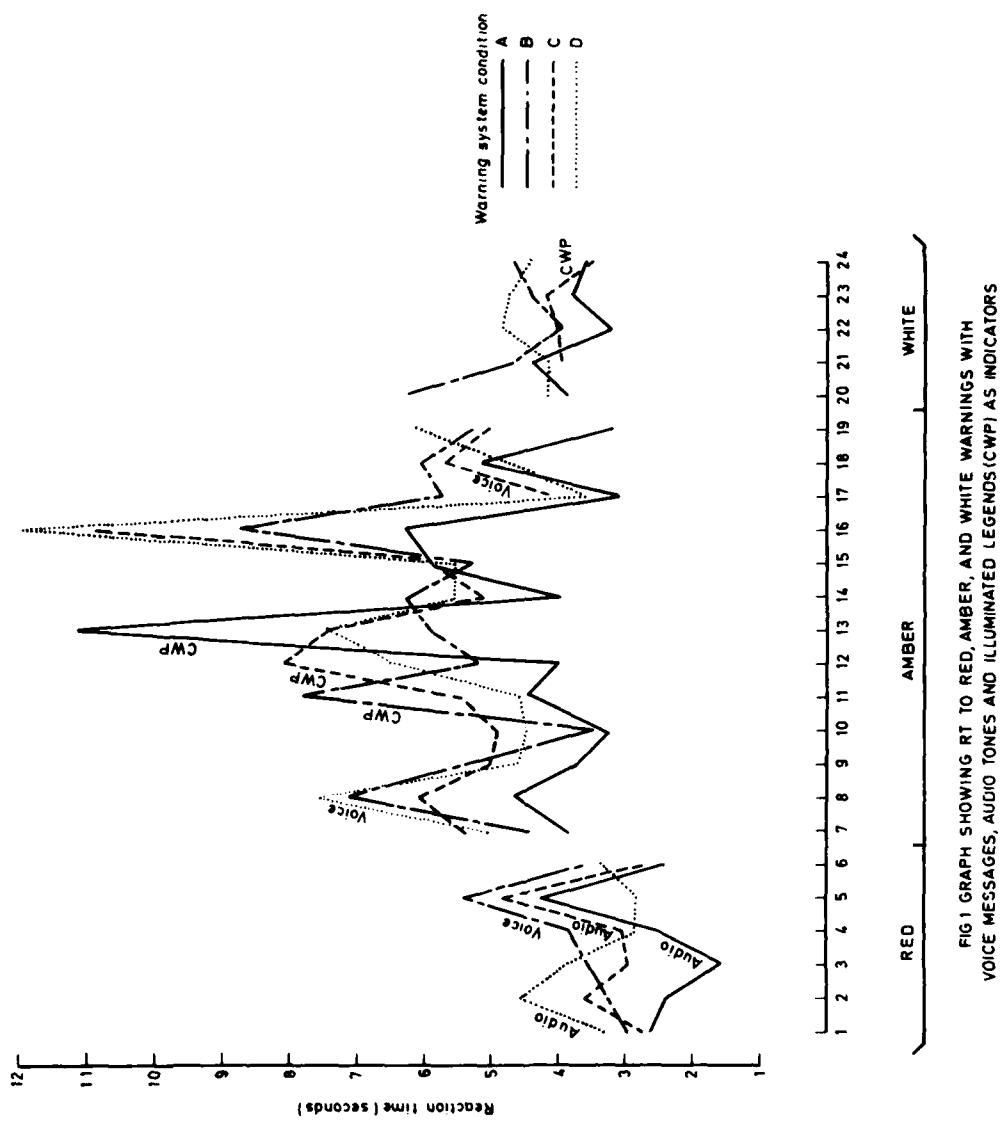
FIG 1 GRAPH SHOWING RT TO RED, AMBER, AND WHITE WARNINGS WITH
VOICE MESSAGES, AUDIO TONES AND ILLUMINATED LEGENDS (CWP) AS INDICATORS
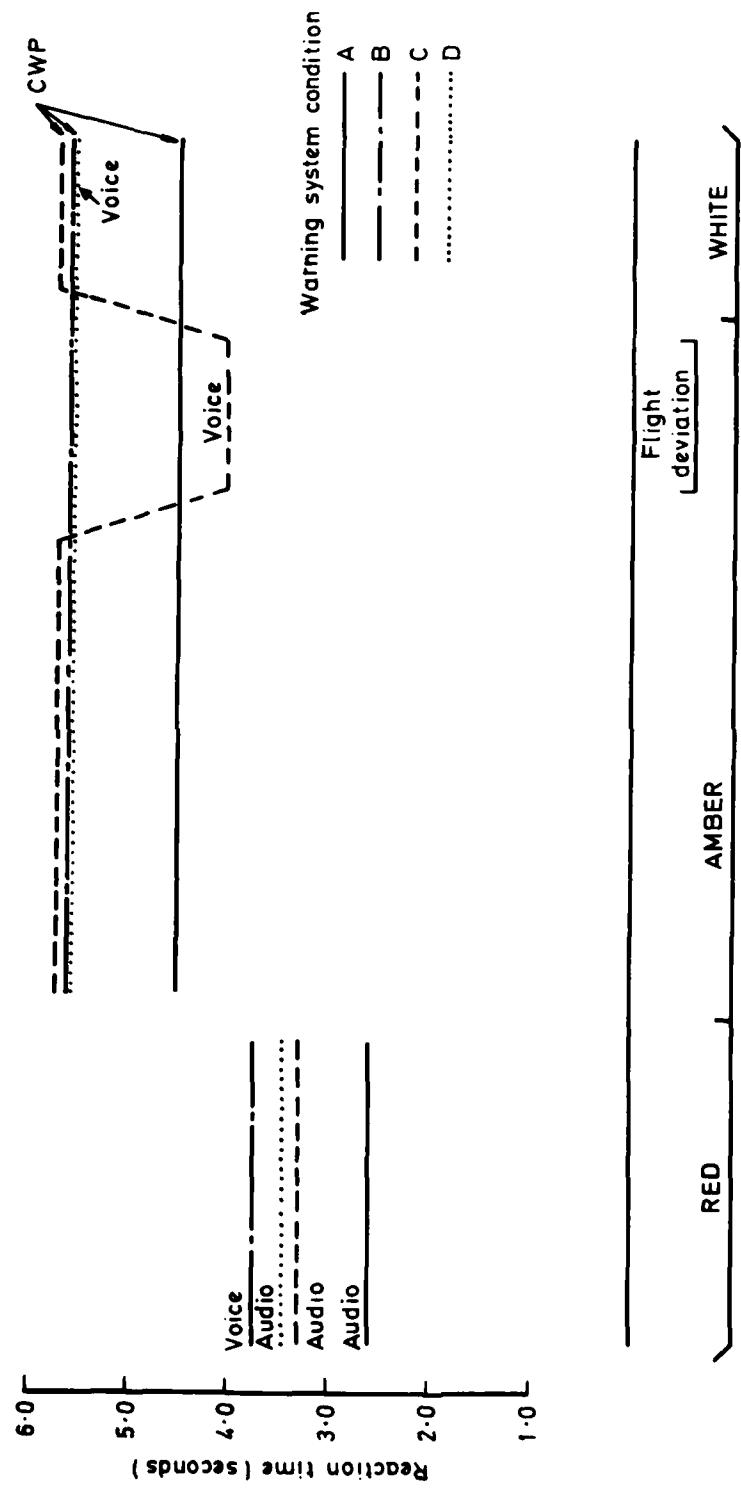
FIG 2 GRAPH SHOWING MEAN RT FOR RED, AMBER AND WHITE WARNINGS
USING VOICE MESSAGES, AND AUDIO TONES AND ILLUMINATED LEGENDS (CWP) AS WARNING INDICATORS
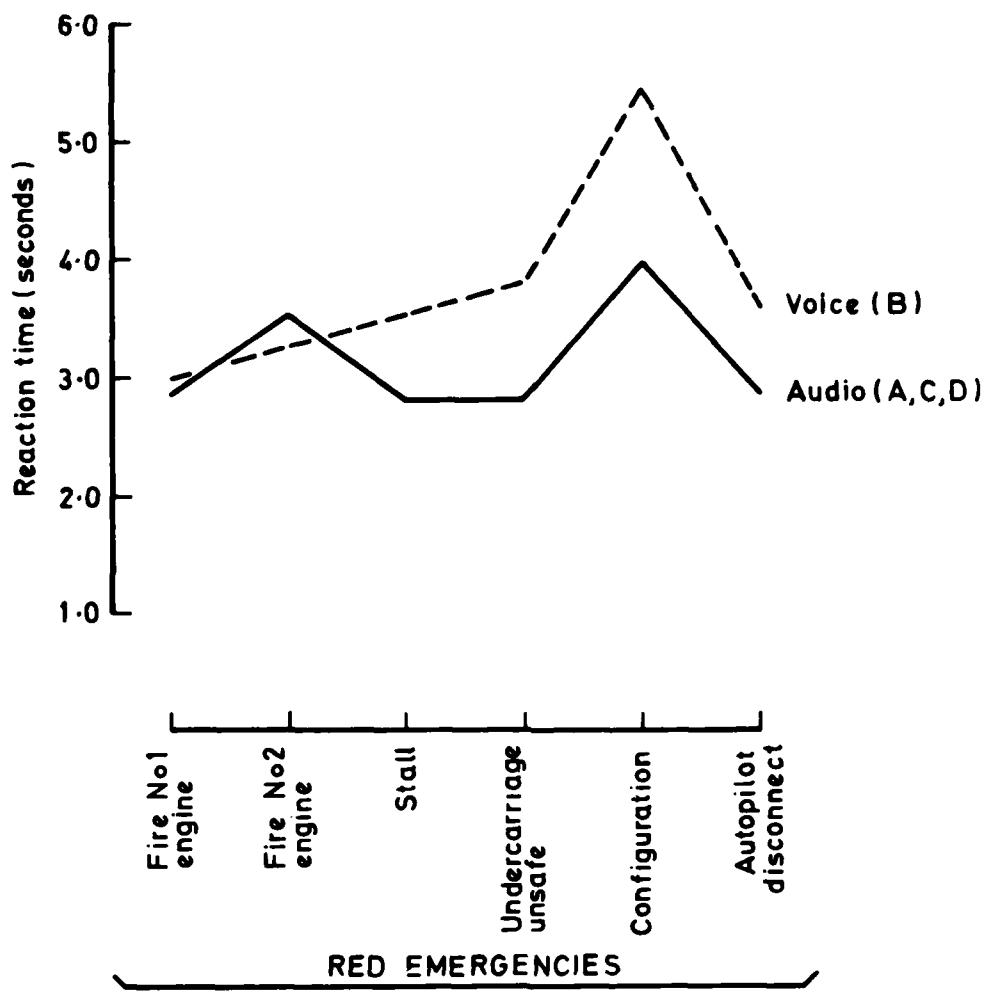
FIG 3 GRAPH SHOWING RT TO WARNINGS FROM
THE RED CATEGORY FOR VOICE MESSAGES AND AUDIO TONES
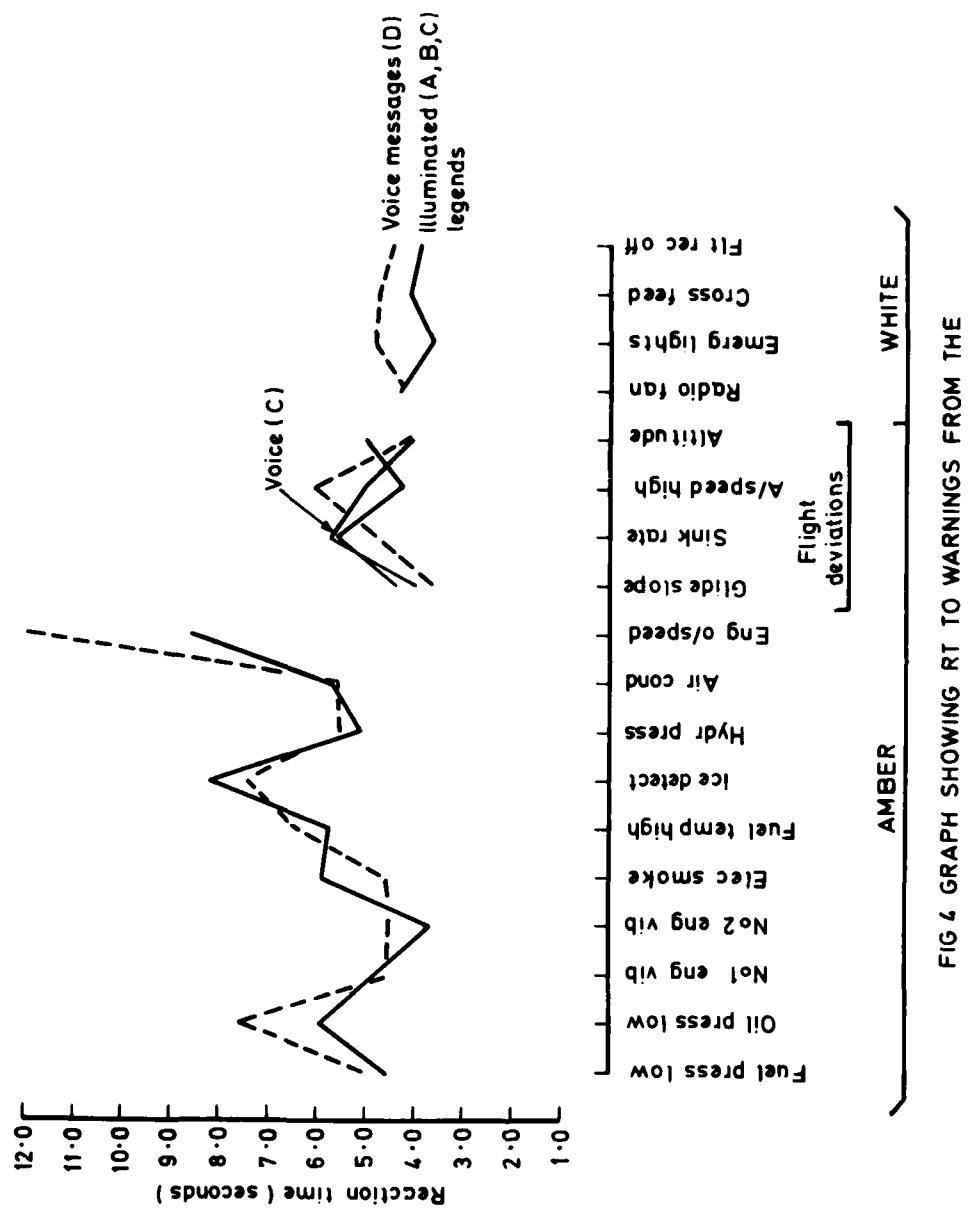
FIG 4 GRAPH SHOWING RT TO WARNINGS FROM THE
AMBER AND WHITE CATEGORIES FOR VOICE MESSAGES AND ILLUMINATED LEGENDS

DISCUSSION

DR. J.D. MOSKO (UNITED STATES)

With regard to the Votrax, as everyone has probably heard, it is not a very good synthesizer, but with regard to the naturalness of the speech, in an informal survey we are doing right now with the Navy Pilots for particular information in a combat environment, for example, giving them air speed, altitude information, their subjective opinion is that they would like the voice to be slightly unnatural. If you sue voice warnings, have your pilots indicated that they might like it less natural, perhaps not over-articulated. In the selective attention literature, you find if you make the spectrum specific or different, they can pick it up a little quicker?

DR. J. WHEALE (UK)

We chose the Votrax voice because it does sound slightly different from the normal voice and because it does so, it has unique attention getting qualities. But the pilots who took part in the experiment didn't appreciate that viewpoint at all. They wanted a perfectly natural sounding voice. They would like a newscaster to read the messages to them, I'm sure.

DR. J.D. MOSKO (UNITED STATES)

Did you take any data on the tracking task, because we are interested in this type of activity because of another test we are developing.

DR. J. WHEALE (UK)

What we did was monitor accuracy on the tracking task for the 15 seconds prior to the onset of the warning and then we monitored accuracy for the 15 seconds subsequent. The data wasn't analyzed in time to include in this paper. It has now been analyzed and it appears that the point at which the performance drops off to a maximum is the same point that the subject is reacting, but given that this point varies according to the length of reaction time, there is no significant difference in the amount of decrement that is produced by the different warning signs. There seems to be a uniform amount of decrement in responding to any warning and you can't differentiate between the warning types. One good thing about the voice warning messages is that they allow you to go from receipt of the message to the action with very little mental effort. But, according to our figures, it is just as easy to go from a very lound sound to response, as it is to go from a nice quiet voice message to response.

DR. R.L. MCKINLEY (UNITED STATES)

Was each one of the voice warning messages identifiable from the other voice warning messages by the first word of the message? I believe that when the US implemented an analog voice warning system on the B-58 each of the voice warning messages was identifiable from the others by the initial word of the warning messages.

DR. J. WHEALE (UK)

Each voice warning message began with the word 'warning'. However, the first word of the key-word warning phrase was unique following the recommendation that came out of the work with the B-58 analog warning system. Current experience has shown that the initial sound of a unique voice message is very important for identification and recognition. For example, the current GPWS warning is structured "Whoop, Whoop, - Pull up, Pull up" and many pilots experienced with the system maintain that all they need to operate effectively is the initial 'whoop' sound.

DR. J.D. MOSKO (UNITED STATES)

Do your subjects prefer natural or less natural speech?

Have you or are you planning experiments to look at more complex decision making tests?

DR. J. WHEALE (UK)

The pilots who took part in the experiment said they could understand the votrax voice warning messages well but that they would have preferred a natural sounding voice with a neutral pattern of information. The above was the majority viewpoint, a minority of pilots viewed the unusual quality of the votrax voice as advantageous with respect to discriminability and attention getting.

We are planning to complete full-mission simulations with experimental warning systems. This will allow us to gauge better the effect of different warning types on complex, realistic flight control tasks.

ASSESSING THE EFFECTIVENESS OF AUDITORY WARNINGS

Peter A. Wilkins,
Audiology and Human Effects Group,
Institute of Sound and Vibration Research,
The University,
Southampton SO9 5NH,
United Kingdom.

SUMMARY

This paper reviews the methods which have been employed to assess the effectiveness of warning sounds. It is emphasised that it is necessary to investigate not only the audibility of the sound, but also its attention demand and recognition. Two laboratory experiments and a field study are described which found that whilst inattention need not necessarily impair the perception of an auditory warning, the combination of inattention and the need to recognise the warning may result in failures in the perception of sounds which can be heard and recognised when listened for deliberately. The results indicated that to be effective a warning sound should be distinct from both the ambient noise and other non-simultaneous discrete sounds present.

These findings imply that when assessing the perception of auditory warnings it is important to consider the acoustic characteristics of both the warning sound and its environment. Whilst it may be possible to predict the effectiveness of some warning sounds, it is recommended that in critical applications their performance should be assessed in field trials using conditions similar to those employed in the present field study.

1. INTRODUCTION

Audio signals are used in a wide range of situations to convey the information of imminent danger. In industry and commerce auditory warning devices are used as alarms for fire, emergency evacuation of a building, burglary, machinery and process faults and the movement of machinery and equipment. In aviation audio warnings are used to inform those in the cockpit of a variety of faults and potentially dangerous conditions, such as fire, ground proximity, faults in electrical and hydraulic systems and aspects of the aircraft's configuration.

The auditory modality is chosen for these applications for various reasons which include its invariance under all lighting conditions, the ability of sound to travel around corners and the generally omni-directional characteristics of the hearing system. In transportation systems in particular there is the additional factor that the operator's visual channel is often already heavily loaded, so that it may be advantageous to use auditory instead of visual signalling.

Inspite of the widespread use of warning sounds, there have been relatively few attempts to determine experimentally their effectiveness under realistic conditions. The term "effectiveness" is used here to describe the probability that the warning will be perceived under typical conditions where its occurrence may be largely unexpected and it may have to be recognised amongst a variety of other sounds. In some instances the time to respond to the warning may be of importance, and this aspect will also be considered briefly.

The dearth of valid empirical data has led to the selection and specification of warning sounds largely on the basis of intuitive judgements. For instance, a report by the General Electric Company (1) derived a criterion which has come to be widely accepted for the intensity of a warning sound in noise on the basis of making the sound sufficiently annoying, whilst Corliss and Jones (2) have suggested a criterion based on the perceived loudness (as distinct from intensity) of the sound relative to the loudness of normal conversational speech.

In this paper the basis of assessing the perception of warning sounds is reviewed and the results of experimental studies are summarised to indicate the principal factors which govern their effectiveness. Particular reference is made in section 3 to a series of experiments conducted at the ISVR which have formed part of a larger programme investigating the effects of wearing hearing protection on the perception of industrial warning sounds (3). Included in this research have been not only those sounds which are emitted intentionally to provide a warning, but also incidental warning sounds which occur in conjunction with imminent danger from machinery or processes. General industrial examples of this latter category include the sound of the engine of an approaching fork-lift truck, the "ringing of the keys" which warns drop-forgers of a loose die (4), and the "roof talk" which occurs in coal mines prior to the collapse of a coal-face (5).

2. CRITERIA OF EFFECTIVENESS

A consideration of the processes involved in the perception of a warning sound identifies the three components shown in the conceptual model in Fig. 1, which intervene between acoustic stimulation and subjective response. The three components represent the reasons why the sound may, or may not, be perceived as a warning. The first component, the audibility of the sound, provides a baseline measure of whether the sound can be heard in the presence of the masking noise when it is listened for deliberately. The

<u>attention demand</u> of the sound determines the ability of the sound to attract a person's attention and <u>therefore be consciously</u> perceived when its occurrence is largely unexpected. The third component, the <u>recognition</u> of the sound as a warning, requires that it can be discriminated from other sounds and that it conveys the meaning of imminent danger.

After the perception of the sound as a warning, an additional process of evaluation by the subject leads to a decision as to what response if any should be made to the warning. This evaluation will be influenced by the context in which the sound is heard and a host of psychological and environmental factors. The credibility of warnings has been discussed by Janis (6) and Latane and Darley (7) and will not be considered further here. The possible criteria of effectiveness of auditory warnings are discussed in the context of the three components of the perceptual process in the following sections.

2.1 Audibility

In the general case the audibility of a warning sound will be limited by the masking effect of background noise. Various procedures have been suggested which can predict the masked threshold of a complex sound in the presence of a particular noise spectrum. (8-12). These procedures are based on frequency analyses of the signal and noise using bandwidths which approximate the critical bands of the ear. The audibility of the sound is governed by the frequency band(s) having the highest signal-to-noise ratio(s). A review of these procedures and a comparison of predicted and measured masked thresholds has indicated that in general use the predictions can be accurate to within $\pm$ 3 dB (12).

One-third octave bands provide a reasonable approximation to the critical bandwidths over the range of centre frequencies 0.5 to 10 kHz, so that a one-third octave band analysis of a particular noise spectrum could in principle be used to specify the optimal frequency of a warning sound for it to be audible with a minimum sound intensity. Because of the variability of most noise environments this consideration is in general likely to be of practical importance only for an indication of a range of suitable signal frequencies.

The predictions described above are based on either monaural or diotic conditons, the latter occurring when identical acoustical stimulation is provided to both ears. Improvements in masked thresholds of up to 15 dB can occur for dichotic conditions where the information to the two ears differs, due to the spatial separation of the signal and noise sources, or artificially varying the phases of the signal and noise separately for each ear (13).

In particular situations the audibility of a signal in noise may be governed by the absolute sensitivity of the hearing system. For listening under free-field conditions the relevant absolute thresholds for people with normal hearing (the minimum audible field) are specified in ISO R226:1961. Absolute threshold limitations on the audibility of a signal in noise may become important when the absolute thresholds are elevated due to a temporary or permanent hearing loss. Because of the predominantly high frequency hearing losess associated with presbyacusis and noise-induced deafness it is generally recommended that warning sounds should have their principal components at frequencies below approximately 2kHz.

2.2 Attention Demand

In specifying the intensity required of a sound to act as a reliable warning in a particular noise environment the masked threshold provides a minimum requirement. However, since an auditory threshold is typically defined as the intensity necessary for a response rate of 50% or 75% (depending on the paradigm employed), this in itself would not provide an entirely reliable warning. Measured detection response rates at different signal-to-noise ratios suggest that a warning sound may have to be 12 dB above its masked threshold to ensure a response of 100% (14).

Over and above this consideration, there remains the question of whether inattention can impair the perception of auditory stimuli. One school of thought argues that hearing is "the sentinel of our senses, always on the alert" (15). Often anecdotal evidence such as a parent's sensitivity to even the faintest baby's cry is reported to emphasise that at least sounds which have high importance to an individual will be reliably perceived (16-17). The alternative school of thought holds that inattention can result in an elevation of the threshold of sounds. It is argued that when a person is involved with a primary task, the demand on attentional capacity may be such that there is insufficient spare capacity (see Fig. 2) for the reliable monitoring of unexpected stimuli (18).

Whilst there is a wide range of experimental work which bears indirectly on the role of attention in the perception of warning sounds (for a review see reference (19)), only relatively few experiments have attempted to measure the possible elevation of threshold due to inattention relative to a measure of the audibility of the sound when it is listened for deliberately. A typical paradigm is the method of paired ascending limits illustrated in Fig. 3 (20-21). An "effective threshold" is measured at random time intervals whilst the subject is continuously performing a loading task. Immediately after this the subject expects a second presentation of the same stimulus to measure the "set threshold". The difference between these two thresholds should provide a measure of the elevation in threshold due to inattention.

The ascending sequence of signal presentations used in this paradigm is appropriate for warning sounds which gradually increase in intensity, as may occur when an emergency vehicle with its siren operating approaches another vehicle. It may however produce an inappropriate elevation of threshold for warning sounds which occur as discrete entities at relatively constant levels. In our own studies we have therefore used the method of constant stimuli, presenting sounds at levels which encompass a range of signal-to-noise ratios in random order (22-23). During an "effective response" condition the subject has been involved with a loading task, a modified version of a television game, and the sounds were presented at randomised time intervals. Prior to and after this condition the audibility of the sound was determined in "detection response" conditions using the same matrix of signals with the sounds being presented in close succession and being expected by the subject. Comparison of the effective and detection response rates at each signal level indicate the change in the probability of perception of the warning

sound due to inattention, as distinct from the single metric elevation of threshold provided by the earlier paradigm.

2.3 Recognition

In most occupational situations warning sounds will occur in a varied acoustic environment, and it is crucial to their effectiveness that they can be recognised amongst the other sounds present. Very little data is available on the recognition of warning sounds under these conditions, but it appears likely that the careful selection of suitable audio warnings can ensure that they will be reliably recognized when they occur non-simultaneously amongst other discrete environmental sounds and are listened for deliberately.

In aircraft cockpits in particular it is also necessary for each individual warning sound to be recognized amongst the set of audio warnings employed. On current generations of civil aircraft there may be between 5 to 9 different audio warnings, each having a different meaning. In a laboratory experiment assessing the ability of airline pilots to recognize the various sounds in use, DuRoss has reported an average error rate of 20% (24). The practical implications of such errors is not however immediately evident, since under operational conditions, provided that a warning is perceived, its exact meaning may subsequently be obtained from visual displays.

Experiments by Patterson and Milroy have indicated that subjects can readily learn between 4 and 6 of the warnings typically used on the flight decks of civil aircraft (25). It was also found that of the set of ten sounds eventually learnt, after one week approximately 6 of the sounds could be correctly identified. These results therefore support the recommendations of standards such as MIL-STD 1472B that with existing sounds a maximum of four different warnings should be employed (26). With this limitation it is necessary to employ master warning signal systems (27), and preferably to standardize on the few sounds to be employed. One suggested scheme for military aircraft is given by MIL-STD-411D which recommends five sounds to cover the meanings: master warning, bail-out, wheels-up, angle of attack/airspeed/stall, and nuclear radiation danger (28).

3. EXPERIMENTAL STUDIES

A series of experiments which have encompassed the criteria of audibility, attention demand and recognition have been performed as part of a research programme which has investigated the effects of wearing hearing protection on the perception of warning sounds. The relevant results of two laboratory-based experiments and a field study are summarised here and compared with other available data.

3.1 Attention demand - Experiment 1 (22)

This experiment investigated the effects of temporal uncertainty alone and in conjunction with a loading task on the response rate to a siren warning sound presented in a background noise of 75 dB(C). One-third octave band spectra of the noise, the siren sound and the calculated signal-to-noise ratios are shown in Fig. 4. In each session the signal was presented four times at each of five different levels in a random order and at random time intervals (mean ISI = 93s, range 20-160s). The five signal levels L1 to L5 were at 5 dB intervals, and were selected so that the masked threshold of the signal was midway between levels L2 and L3. The experiment was conducted in an anechoic room with the signals and noise presented from a loudspeaker located directly in front of the subject.

Twelve subjects with hearing levels better than 20 dB re ISO 389-1975 in the range 0.5 to 6 kHz in both ears participated in the experiment. Their mean age was 20 years with a range of 19 to 23 years. Each subject attempted one session under vigil conditions, and another on a separate day with the same temporal uncertainty but with the subject performing the loading task. The display of the loading task, a modified version of a television tennis game, is shown in Fig. 5. The subject controlled the "bat" on the left and attempted to direct the "ball" through the moving "hole" on the right. The subjects were encouraged in their performance at the task by continuous feedback of the points score by the marker at the bottom of the screen, and at both the task and responding to the warning sound by a financial bonus awarded at the end of the experiment on the basis of their overall performance.

In addition to the effective response (ER) condition which made up the main part of each session, the subject's ability to detect the warning sound was measured prior to and after the 31 minute ER condition in the detection response conditions, DR1 and DR2 respectively. The number of responses for these three conditions as a percentage of the total number of signal presentations are shown separately for the vigil and task sessions in Fig. 6 as a function of the signal level. The curves show the typical sigmoid form of psychometric functions, rising from a chance rate at level L1 to 100% response rate at L5. For both the Vigil and the Task condition only small differences are evident between the effective response and detection response conditions.

Statistical analysis indicated that only the difference between response conditions at level L2 was significant (p<0.01 by Tukey's test). The results therefore show that the conditions of inattention in the ER condition may have slightly reduced the response rate for the signal at sub-threshold levels, but that at supra-threshold intensities there was no evidence of an effect of inattention on the perception of the siren.

The absence of any significant differences in response rates between the Vigil and Task conditions is in agreement with the results of Emson (29). In the context of a capacity model of attention (18), the absence of any additional effect of the loading task suggests that with the demand on attentional capacity associated with typical psychomotor tasks there is normally sufficient spare capacity available for the perception of important auditory stimuli.

The absence of any effect of inattention on perception in this experiment is in contrast with the elevation of threshold reported in two similar studies (20, 30). This may be due to methodological

differences including the use of a method of ascending limits and inappropriate motivation of the sub-
jects in both these studies, which could have produced the elevation of threshold as an artefact.   A
further difference is the incorporation of an element of recognition in the two previous studies since
several warning sounds were tested in the one session but the subjects were uncertain as to which would
occur next.

It can be concluded from experiment 1 that inattention due to temporal uncertainty alone or in con-
junction with a loading task will not necessarily reduce the probability of the perception of a warning
sound.

3.2 Attention demand and recognition - Experiment 2 (23)

A second experiment has been performed which provided an additional requirement during the effective
response condition that the warning sound must be recognised amongst other discrete non-simultaneous
sounds. This provides a more realistic simulation of most occupational situations where the occurrence
of the warning sound is largely unexpected,and to be perceived   must be filtered out from a large
number of irrelevant sounds which do not reach conscious attention.

The warning and irrelevant sounds were selected such that the warning sounds could be recognised
when listened for deliberately. Thus in terms of the conceptual model shown in Fig. 1 the principal in-
terest is in the possible interaction between the recognition and attention demand components.

The design and conditions of the experiment were essentially the same as those in experiment 1.
However, in this case sixteen (normally hearing) subjects participated in the experiment (mean age 22
years, range 21 to 26 years). A set of four discrete machinery sounds was selected to provide a typical
background of meaningful stimuli amongst which the siren would occur. The sounds varied in their
spectral content over their nominal duration of 5s and so were comparable in this respect to the siren.
It is however evident from the typical power spectra of the noise and the five sounds shown in Fig. 7
that the tonal character of the siren is distinctive from the ambient noise and the other four machinery
sounds. In the effective response condition the five sounds were presented at one of the five levels L1
to L5 in random order and with random time intervals. The intensities of the sounds were increased by
5 dB relative to the corresponding levels in experiment 1 so that the masked thresholds of the sounds
were midway between levels L1 and L2 in this experiment. The mean inter-stimulus interval of 19s
(range 8s to 28s) gave rise to a mean inter-signal interval of 95s. In addition to the effective and
detection response conditions (ER, DR1 and DR2 respectively), the subjects' ability to recognize the
warning sound was measured when it was listened for deliberately amongst the other irrelevant sounds in
a recognition response condition RR.

In one session the siren acted as the warning sound with the four machinery sounds as irrelevant
stimuli. The results for this condition can then be compared directly with those from experiment 1. In
the other session on a separate day the grinder sound acted as the warning amongst the four remaining
sounds in the set, to assess the relative effectiveness of a typical incidental warning sound. In both
sessions subjects performed the loading task to provide inattention additional to the temporal uncertain-
ty associated with the occurrence of the warnings.

The proportion of responses for the four response conditions (ER, RR, DR1 and DR2) are shown
separately for the two warning sounds as a function of the signal level in Fig. 8. It is immediately
evident that for the siren there were no large differences between any of the response conditions. In
fact at levels L4 and L5 the siren was detected, recognised and perceived in the effective response
condition with 100% reliability. It can therefore be concluded that the additional cognitive load
relative to experiment 1 of having to recognise the siren did not impair its attention demand.

By contrast, there are significant differences between the response conditions for the grinder sound.
Averaged over the levels L2 to L5 the effective response rate is 9% less than the mean detection response
rate DR, where DR = $\frac{1}{2}$(DR1+DR2). Whilst at level L2 this can be more than accounted for by difficulties
in the recognition of the sound per se, it is evident that at levels L3 to L5 this is not the case.
The failures in perception in the effective response condition at these levels must therefore be due to
the combined conditions of inattention and recognition.

3.3 Field study (31)

The majority of studies into the perception of warning sounds have been conducted in a laboratory
setting, primarily due to considerations of convenience and to permit precise control of the variables
involved. In moving from the real world to the laboratory there are however risks that erroneous con-
clusions are made due to inaccurate simulation of relevant factors (32). A field study has therefore
been performed in an attempt to validate the principal findings of the laboratory studies.

The field study was conducted in the press shop of a large factory. The subjects were the press
operators and their perception of the warning sounds was assessed whilst they were performing their
normal work tasks. Each subject was provided with a response button to indicate whenever they heard the
warning, but otherwise continued with their normal work routine.

Two warning sounds which occurred in the press shop were investigated in the study. One was the
horn used by the fork-lift trucks to warn of their approach as they moved around the presses. The
second was the clinking sound of metal components spilling from their container. In addition three
further machinery sounds were tape recorded, the sound of two different presses in operation and the
sound of the overhead crane starting to move. These sounds were mixed onto two master tapes to provide
a format of presentations similar to that employed in experiment 2. The random time intervals between
presentations in the effective response condition were drawn from a rectangular distribution over the
range 10 to 50s giving a mean inter-stimulus interval of 30s, and a mean inter-signal interval of 150s.
The stimuli were presented from a single loud-speaker so that their intensity varied over a range of
approximately 7 dB(A) at the different work stations, whilst a survey of the noise levels at each

station indicated variations in noise level of approximately 10 dB(A) with time at each station and between different stations.

Data were collected on four separate days with four 50 minute blocks of presentations occurring on each day. The effectiveness of each warning sound was assessed separately on two of the days. At any one time a maximum of eight subjects participated in the experiment. However because of the varying work schedules in the press shop it was not possible to obtain balanced sets of subjects and work stations for the two warning sounds. In all 30 subjects participated in the experiment, covering a range of ages from 19 to 59 years (mean 43 years, median 48 years). There was a correspondingly wide variation in the hearing sensitivity of the subjects as measured by pure tone audiometry.

The proportions of responses to the presentation of the two warnings at the levels L3 to L5 for the effective and detection response conditions are shown in Table 1. These percentages are not based on exactly the same number of presentations, there being between 64 and 74 presentations of each sound at each level in the effective response condition and between 42 and 44 for the detection response condition.

Table 1. The proportion of responses (%) to presentations of the two warning sounds for the effective and the detection response conditions in the field study, as a function of the signal level.

Level		L3	L4	L5
Sound	Response Condition			
Horn	Effective response	37	83	91
	Detection response	57	84	89
Clink	Effective response	23	78	83
	Detection response	21	62	67

Apart from the increasing response rate with increasing signal level, none of the differences in the table are statistically significant. The failure of the detection response rates for both sounds to reach 100% at even the highest presentation level suggests an unwillingness on the part of the subjects to respond to all the sounds heard in this condition. It was observed that at level L5 the sounds were clearly audible, but due to the nature of the work task it appeared that the subjects were not willing to respond repeatedly to sounds presented in close succession. In addition, it is likely that the lower detection response rates for the clink sound can in part be attributed to the inability of those subjects with a substantial hearing loss to detect its predominant high frequency components, and in part to the greater difficulty in recognising it amongst the other naturally occurring machinery sounds (31).

Notwithstanding these limitations, it is evident that whilst effective and detection response rates to the horn are similar at levels L4 and L5, at level L3 there is a 20% reduction in the response rate due to the conditions of inattention and the recognition requirement. This corresponds with the similar trend for the siren sound at the near threshold level of L2 evident in Fig. 6.

As the intensities of the two sounds were selected prior to the experiment to ensure approximately equal audibility at the corresponding presentation levels, a comparison of the effective response rates for the two sounds would suggest that under the conditions investigated the horn was a more effective warning than the clink sound. This could in part be due to the reduced audibility of the clink sound for those subjects with a hearing loss, but probably also includes an effect due to the greater difficulty in recognising the clink amongst the other real and artificially generated machinery sounds.

It is therefore to be concluded that the results of the field study are not in complete agreement with the laboratory-based studies. In particular the earlier observation that a distinctive intentional warning sound would act as an entirely reliable warning was not corroborated by the 91% effective response rate of the horn at level L5. In addition the maximum effective response rate of 83% for the clink sound suggests that considerably larger effects than were observed in the laboratory-based experiments may occur under realistic conditions. There is therefore a need for further laboratory and field experimentation to examine the relative importance of the various differences in the conditions of the present studies.

4. DISCUSSION

The results of experiment 2 provide clear evidence for the first time that apart from their audibility, there can be differences in the effectiveness of different warning sounds. It is possible that the grinder was not entirely reliable as a warning because it lacked the psychological associations with danger of the siren. It was however hypothesised that the differences between the effectiveness of the two sounds can be explained in terms of their acoustical characteristics.

For the grinder, it is evident from Fig. 7 that its generally broad-band character makes it similar to the other three machinery sounds and the ambient noise, whilst the siren was a highly distinctive sound. From the data of the present experiments it is not however evident which of these two contrast factors would be the predominant influences on the effectiveness of the grinder sound. A further experiment has therefore been conducted which indicated that both factors can be important, but that for the stimuli investigated the contrast relative to the other irrelevant sounds was more important than the contrast relative to the ambient noise (33). The variation in effectiveness of the sounds as warnings in this latter experiment due to manipulation of these acoustical factors supports the hypothesis that the relative effectiveness of warning sounds can in the first instance be attributed to physical rather than purely

psychological factors. In particular the results indicate that the attention demand of a warning sound may depend as much on the environment in which it occurs as on its own acoustical characteristics.

Because of the practical constraints in performing the field study described in section 4.3, its results can only provide a general comparison with the laboratory-based experimentation. Whilst there was some indication of a difference in effectiveness of the intentional and the incidental warning sounds in the field study, there were sufficient differences with the results of the laboratory experiments to suggest the need for further comparative studies. It is intended in future experimentation to overlap field and laboratory studies so that the relative importance of variables such as the hearing level of the subjects, their work activities, the noise environment and the inter-signal interval can be assessed.

Until a reliable means of predicting the effectiveness of auditory warnings can be derived from suitable laboratory experimentation, there will be a need for assessments of effectiveness under actual working conditions. A recent German standard has recommended a procedure for the assessment of industrial warning sounds which requires conditions similar to these employed in the present field study (34). Whilst flight simulators have been used in some studies assessing the effectiveness of aviation warnings (35), there is no equivalent standard for their assessments under actual flight conditions.

Although there have been numerous claims that some warning sounds are more effective than others (36, 37) when account is taken of their audibility it appears that there are in general no measurable differences (21, 30, 38). Whilst the results of the present experiment do not provide any direct evidence of differences between types of intentional warning sounds, they suggest that if the sound is not sufficiently distinct, as typically occurs with incidental warning sounds, it may not act as an entirely reliable warning. Of particular relevance to the conditions on flight-decks where there are several warning sounds was the finding in the additional experiment (33) that the perception of even a tonal warning sound can be impaired by the need to recognise it amongst other tonal sounds. This suggests that even with a small number of audio warnings in a cockpit, there may be failures in perception if the sounds are not sufficiently distinct.

The difference between the effective and detection response rates for the grinder in experiment 2 is relatively large at level L4 and exists at the highest level investigated (L5). This effect would not however be revealed by the single figure elevation of threshold metric used in the earlier studies (21,30). It also suggests that the generally accepted requirement that a warning sound should be at least 15 dB above its masked threshold (1, 27) may not in fact ensure an entirely reliable warning.

Response time measures in the effective response condition for the two sounds in experiment 2 showed a significantly larger increase relative to the latencies in the detection response condition for the grinder than the siren (mean increases over levels L2 to L5 of 0.51s and 0.18s respectively). However a regression between this measure and the corresponding difference between response rates accounted for only 7% of the total variance in the latter, thus indicating that any relationship between the two measures is not of sufficient strength to permit a predictive relationship (12). There is therefore no basis in this study for the use made in other studies of response latencies to assess the effectiveness of warning sounds (39, 40). Given that differences in response times between warning sounds will only be fractions of seconds, there is no a priori basis for the assumption that a shorter response latency indicates a more effective warning.

Although it has not been investigated in the present experiment, it has been presumed that there is an upper limit for the intensity of a warning sound. This may be on the basis of preventing annoyance or hearing damage, or of actually disrupting the response to the warning by causing startle or panic, and unnecessarily interfering with other communications. Two suggestions of such an upper limit have been the masked threshold plus 25 dB (11), and the intensity mid-way between the masked threshold and 110 dB (27).

5. CONCLUSION

This review of the methods employed in assessing the effectiveness of warning sounds has emphasised the necessity of accounting for all aspects of the perceptual processes involved. In particular it has been shown that it is necessary to assess not only the audibility of the sounds, but also their ability to demand a person's attention and be recognised so as to communicate their meaning of impending danger reliably. It has been found in laboratory experiments that the need to recognise the auditory warning amongst other non-simultaneous irrelevant sound may impair its perception when its occurrence is largely unexpected. This finding has emphasised the need to simulate the acoustic environment realistically in future laboratory experiments. The results of a field study have indicated the necessity of further investigating the role of the hearing level and work activity of the subjects, and other aspects of the acoustic environment.

The development of more fully validated procedures for assessing and predicting the effectiveness of auditory warnings would be of considerable benefit in standardisation exercises. Whilst applying in the first instance to the effectiveness of non-speech warnings presented in the auditory mode, similar procedures could be applied to voice warnings, to those in the visual and tactile modes, and to the various combinations of these.

6. REFERENCES

(1) GENERAL ELECTRIC (1963) Improved outdoor warning devices. Report OCD-OS-62-171 Off. Civ. Def., Washington D.C.

(2) CORLISS, E.L.R. and JONES, F.E. (1976) Method for estimating the audibility and effective loudness of sirens and speech in automobiles. J. Acous. Soc. Am. 60, 1126-1131.

(3) WILKINS, P.A. and MARTIN, A.M. (1981) The effects of hearing protection on the perception of warning sounds.In:Proc. Int. Symp. on Personal Hearing Protection in Industry. Croom Helm, London.

(4)   CHRIST, E. (1976) Individual hearing protection and accident hazard (German).   Die
          Berufsgenossenschaft, No. 12, 474-480.

(5)   MICHAEL, P.L., SAPERSTEIN, L.W. and PROUT, J.H. (1972) Aspects of noise generation and hearing
          protection in underground coal mines. U.S. Bur. of Mines Final Rep. Grant No. G0122004.

(6)   JANIS, I.L. (1962) Psychological effects of warnings. In: Man and Society in Disaster.   Eds.
          G.W. Baker and D.W. Chapman. Basic Books, New York.

(7)   LATANE, B. and DARLEY, J. (1970) The Unresponsive Bystander: Why Doesn't He Help? Appleton-Century
          Crofts, New York.

(8)   LICKLIDER, J.C.R. (1961) Audio warning signals for Air Force weapons systems. WADD Tech. Rep.60-814,
          Wright Patterson Air Force Base, Ohio.

(9)   SCHOLL, H. (1963) Uber ein objectives vergahren zur emittlung von hörschwellen und mithörschwellen.
          Freqeunz 17, 125-133.

(10) FIDELL, S., PEARSONS, K.S. and BENNETT, R.L. (1974) Prediction of aural detectability of noise signals.
          Human Factors 16, 373-383.

(11) PATTERSON, R.D. and MILROY, R. (1979) Existing and recommended levels for auditory warnings on civil
          aircraft. Contract Report 7D/S/0142, MRC Appl. Psychol. Unit, Cambridge.

(12) WILKINS, P.A. (1980) The effects of wearing hearing protection on the perception of warning sounds.
          Ph.D. Thesis, The University, Southampton.

(13) WEBSTER, F.A. (1951) The influence of interaural phase on masked thresholds. I The role of interaural
          time deviation. J. Acoust. Soc. Am. 23, 452-462.

(14) WILKINS, P.A. and MARTIN, A.M. (1981) The role of acoustical characteristics in the perception of
          warning sounds and the effects of wearing hearing protection. Manuscript in preparation.

(15) DAVIS, H. (1970) Preface to: S.S. Stevens and F. Warshofsky, Sound and Hearing. Time-Life Books,
          Netherlands.

(16) SEMOTAN, J. and SEMOTANOVA, M. (1969) Startle and other human responses to noise. J. Sound Vib. 10,
          480-489.

(17) NEISSER, U. (1967) Cognitive psychology. Appleton-Century Crofts, New York.

(18) KAHNEMAN, D. (1973) Attention and effort. Prentice Hill, Englewood Cliffs, New Jersey.

(19) WILKINS, P.A. and MARTIN, A.M. (1978) The effect of hearing protectors on the perception of warning
          and indicator sounds - A general review.   Tech. Rep. 98, ISVR, The University, Southampton.

(20) KREEZER, G.L. (1958) A threshold method measuring the attention-demand value of stimuli. Am. J.
          Psychol. 71, 111-122.

(21) KREEZER, G.L. (1959) Attention value of audio and visual warning signals.   WADC Tech. Rep. 58-521,
          Wright Patterson Air Force Base, Ohio.

(22) WILKINS, P.A. and MARTIN, A.M. (1978) The role of attention in sound perception.   Proc. Inst. of
          Acoustics 4, 15.01.1-15.01.4.

(23) WILKINS, P.A. and MARTIN, A.M. (1980) The role of recognition in the perception of unattended
          auditory stimuli. In Disorders of Auditory Function III.   Eds. I.G. Taylor and A. Markides,
          Academic Press, London.

(24) DUROSS , S.H. (1978) Civil aircraft warning systems - A survey of pilot opinion within British Airways.
          Tech. Rep. 78056, Royal Aircraft Establishment, Farnborough.

(25) PATTERSON, R.D. and MILROY, R. (1980) Auditory warnings on civil aircraft: the learning and retention
          of warnings. Final Contract Report 7D/S/0142, MRC Appl. Psychol. Unit, Cambridge.

(26) US DEP. DEF. (1974) Human engineering design criteria for military systems, equipment and facilities.
          U.S. Mil. Stand. MIL-STD-1472B.

(27) VEITENGRUBER, J.E., BOUCEK, G.P. and SMITH, W.D. (1977) Aircraft alerting systems criteria study,
          Vols. 1 and 2. Fed. Aviat. Auth. Rep. FAA-Rd-76-222 (1 and 2).

(28) US DEP. DEF. (1974) Military standard aircrew station signals. U.S. Mil. Stand. MIL-STD-411D.

(29) EMSON, I.C. (1962) Audible warnings for fire appliances. Inst. Fire Eng. Q.22, 44-51.

(30) POTTER, R.C., FIDELL, S.A., MYLES, M.M. and KEAST, D.N. (1977) Effectiveness of audible warning
          devices on emergency vehicles. Rpeort No. DOT-TSC-77-38, US Dept. of Transp, Washington D.C.

(31) WILKINS, P.A. (1981) A field study to assess the effects of wearing hearing protectors on the per-
          ception of warning sounds in an industrial environment. Manuscript in preparation.

(32) SMITH, R.L. and LUCACCINI, L.F. (1969) Vigilance research: Its application to industrial problems. Hum. Factors 11, 149-156.

(33) WILKINS, P.A. and MARTIN, A.M. (1979) The role of recognition and sound characteristics in the perception of unattended auditory stimuli. Paper presented at the Short Papers Meeting of the British Society of Audiology, Cambridge.

(34) DTSCH. INST. FÜR NORMUNG (1977) Acoustic signals: signals of danger for work places. DIN 33 404.

(35) KEMMERLING, P. and GEISELHART, R. (1969) A comparison of voice and tone warning systems as a function of task loading. ASD-TR-69-104, Wright-Patterson Air Force Base, Ohio.

(36) SPENCER, J. (1953) A comparison of an auditory warning system with a controlled visual warning system for use in aircraft. Unpub. Rep. RAF Inst. of Aviat. Med., Farnborough.

(37) McCLELLAND, I.L. (1980) Audible warning signals for police vehicles. Appl. Ergon. 11, 165-170.

(38) WHATSON, M.E.C. (1979) The relative effectiveness of warning sounds. Unpub. B.Sc. Thesis, The University, Southampton.

(39) FIDELL, S. (1978) Effectiveness of audible warning signals for emergency vehicles. Hum. Factors, 20, 19-26.

(40) SIMPSON, C.A. and WILLIAMS, D.H. (1980) Response time effects of alerting tone and semantic context for synthesized voice cockpit warnings. Hum. Factors 22, 319-330.
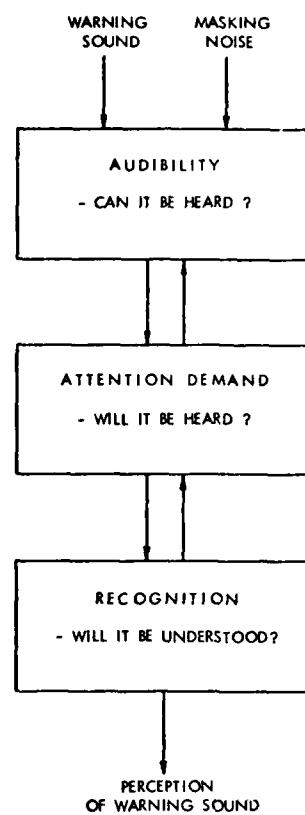
FIG. 1   CONCEPTUAL MODEL OF THE PERCEPTION OF A WARNING SOUND.
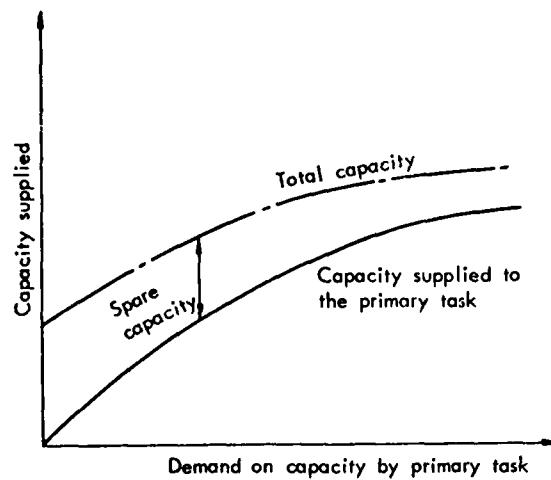
FIG. 2   THE RELATIONSHIP BETWEEN SPARE CAPACITY AND DEMAND ON CAPACITY BY
A PRIMARY TASK (18).
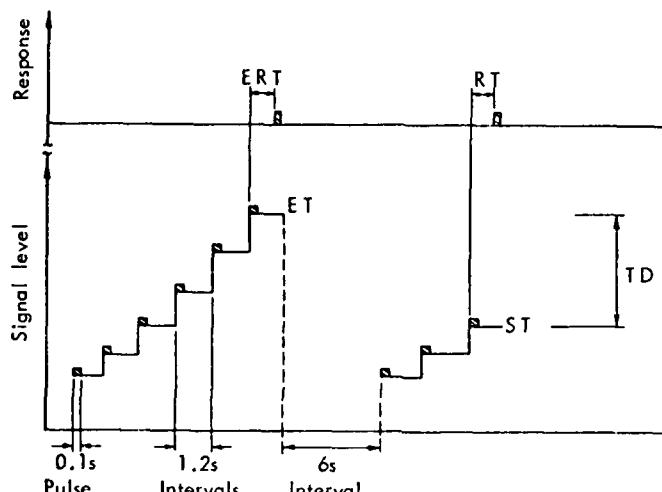
Threshold difference = effective threshold - set threshold
          TD   ET   ST

Reaction time difference = effective reaction time - set reaction time
          RTD   ERT   SRT

FIG. 3   THE TIME HISTORY OF EVENTS IN THE METHOD OF PAIRED ASCENDING
          LIMITS USED BY KREEZER, 1959 (21)

(b)

(a)
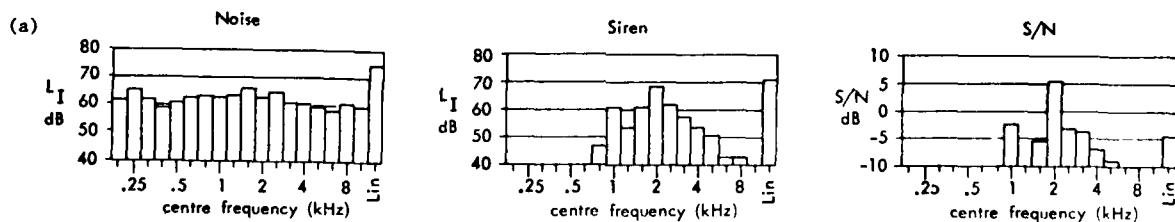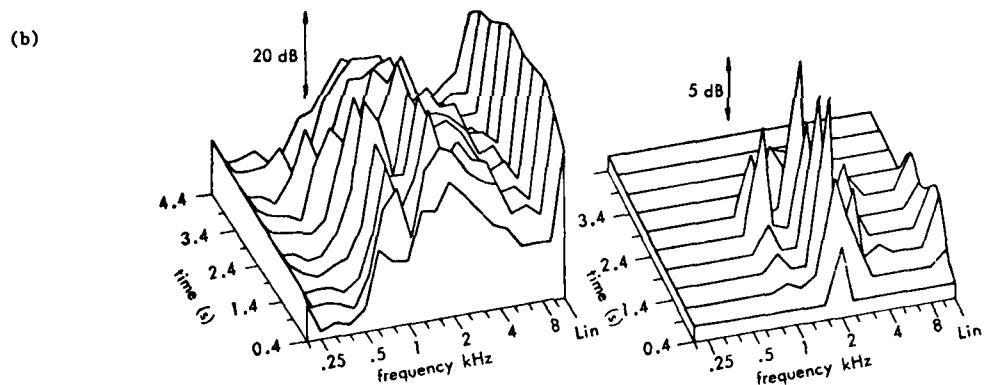
FIG. 4   ONE-THIRD OCTAVE BAND SPECTRA OF THE SIGNAL, NOISE AND SIGNAL-TO-NOISE
          RATIO; THE S/N APPROXIMATELY INDICATES THE COMPONENTS OF THE SIREN
          AUDIBLE ABOVE THE NOISE WITH THE SIREN AT LEVEL L5.

          (a) L_{eq} - TYPE INTEGRATED ENERGY MEASURE (L_I)

          (b) THREE-DIMENSIONAL DISPLAYS INDICATING THE TEMPORAL VARIATION OF THE
              ONE THIRD OCTAVE BAND SPECTRUM; INTEGRATION PERIODS 0.1s WITH 0.5s
              BETWEEN SPECTRA.

Ball horizontal speed 240mm s^{-1}
Hole speed   200mm s^{-1}

TDC Top dead centre
BDC Bottom dead centre

All dimensions in mm
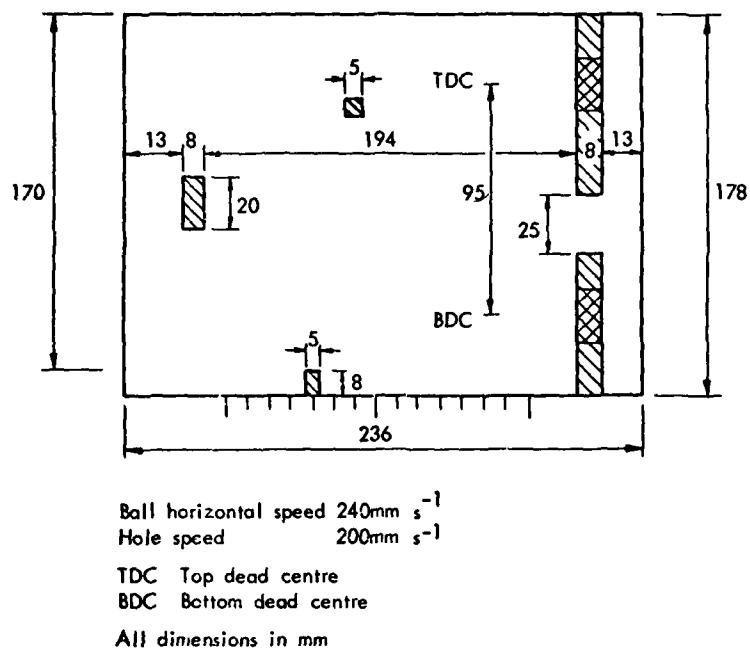
FIG. 5   DIMENSIONS OF THE LOADING TASK DISPLAY.

x------ DR1   pre-test detection response rate
o——— ER   effective response rate
+——— DR2   post-test detection response rate
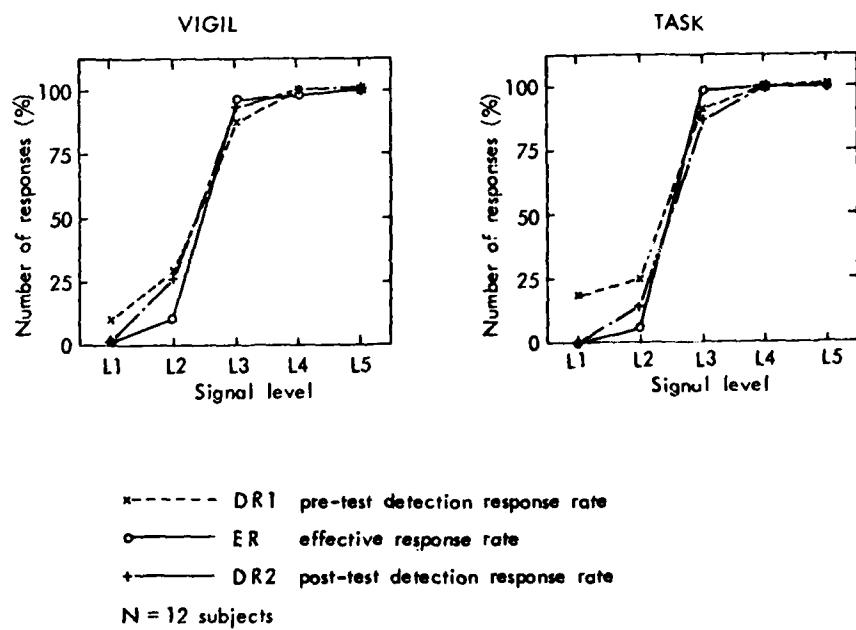
N = 12 subjects

FIG. 6   NUMBER OF RESPONSES AVERAGED ACROSS SUBJECTS AS A FUNCTION OF THE
SIGNAL LEVEL FOR THE TWO TASK CONDITIONS IN EXPERIMENT 1.
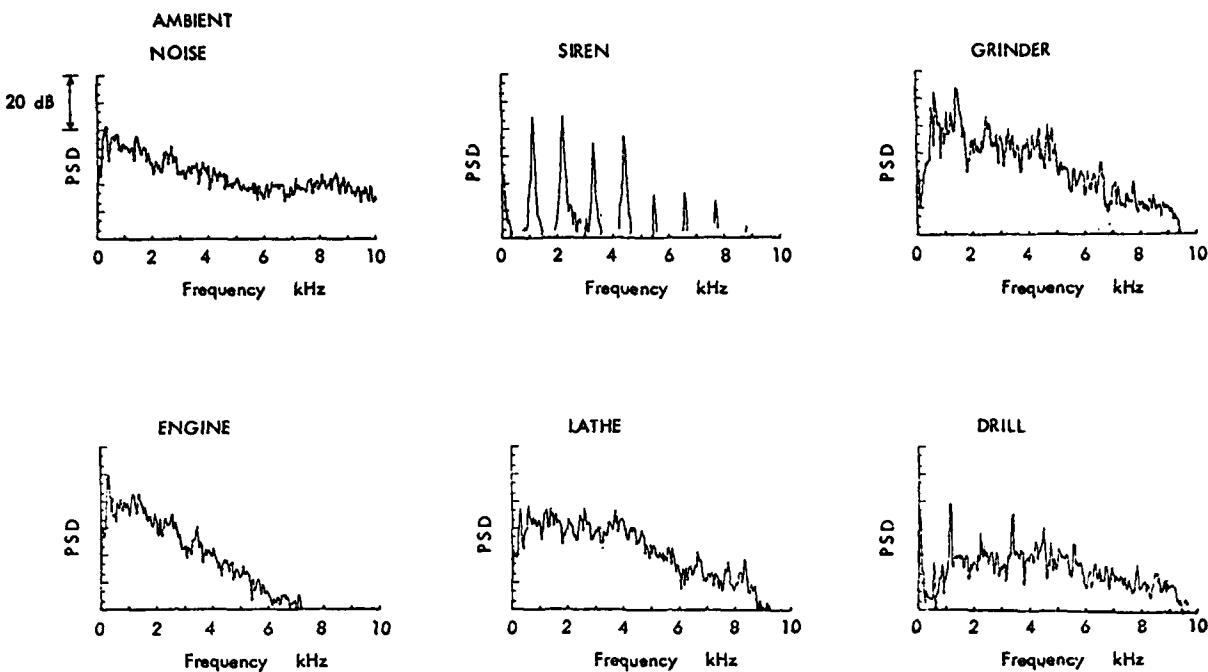
FIG. 7   TYPICAL POWER SPECTRAL DENSITIES (PSD) OF THE FIVE SOUNDS
EMPLOYED IN EXPERIMENT 2.

x----- DR1   pre-test detection response rate
o———— ER   effective response rate
+—·— DR2   post-test detection response rate
△———— RR   recognition response rate
N = 16 subjects
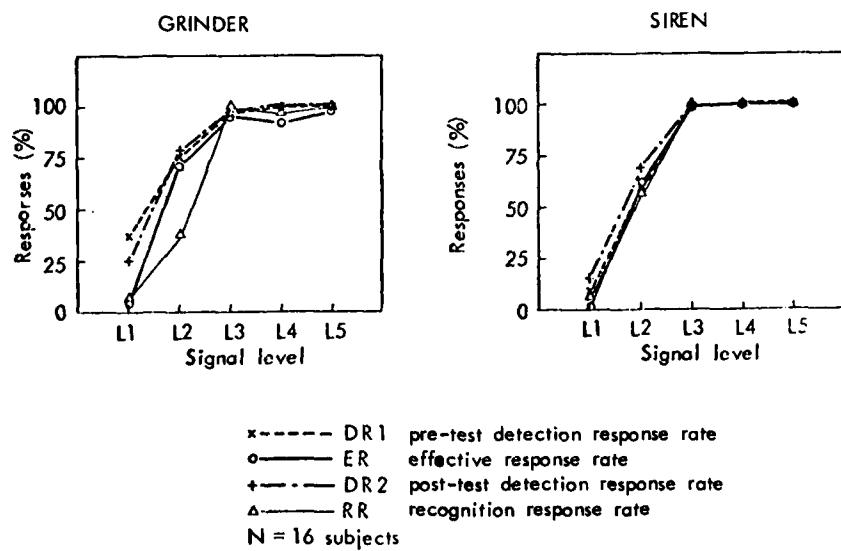
FIG. 8   NUMBER OF RESPONSES AVERAGED ACROSS SUBJECTS AS A FUNCTION
OF THE SIGNAL LEVEL FOR THE TWO WARNING SOUNDS IN EXPERIMENT 2.

DISCUSSION

DR. R.L. MCKINLEY (UNITED STATES)

Could you detail the levels of L1 through L5 please?

Is it correct that L1 is then 5dB above masked threshold?

What is the precise level of L3?

DR. P.A. WILKINS (AUSTRALIA/UK)

Throughout my studies the masked threshold is used as the baseline measure. The levels L1 to L5 were in each case selected to span across the masked threshold at 5dB intervals to encompass the region in which partial masking is an important factor. In the first experiment the masked threshold was midway between levels L2 and L3, in the second midway between levels L1 and L2 (see Figures 6 and 8 with a 50% response criterion).

VOICE WARNING SYSTEMS: SOME EXPERIMENTAL EVIDENCE CONCERNING APPLICATION

Michael Reinecke

GAF Institute of Aviation Medicine, 8072 Manching, FRG

SUMMARY

Two experiments with voice warning systems (VWS), one in a helicopter UH-1D and the other one in a F 104 flight simulator are described.

In the first experiment recognition times to identify simulated failures were measured in cruise - and low level - flights with 5 pilots. It could be proved that voice warnings compared to light warnings do reduce recognition time. This is especially true during low level flight, and when only precise warning texts are used.

In the second experiment the interaction of voice warnings and radio communication was investigated. 11 pilots had to do a navigation flight and to react with correct emergency procedures when failures were introduced. Reaction times suggest, that additional light warnings tend to slow down pilots reactions. Our findings stress the possibility that the pilot might become overloaded when voice warnings do occur while radio communication is going on.

INTRODUCTION

When voice warning systems were first introduced some 20 years ago, scientists and aircraft designers believed they could easily combine two things:

First, it was known that pilots did react faster on audio warnings (such as the sound of a siren or a bell) than on warning lights, especially when their visual channel was loaded with other tasks.

Second, the introduction of a recorded voice meant that specific information could be given now on a theoretically infinite number of different emergency situations.

There were of course some technical limitations in the first generation of voice warning systems. This was due to the fact that only magnetic tape could be used for storing of the voice warnings. Thus the number of messages was limited, while the electro-mechanical playback system caused some time delay and was prone to mechanical failure.

Today these problems have been overcome. Engineers and researchers are enthousiastic about digitally stored voice and synthetic speech. After 10 years of relative calmness pilots must be prepared for voice warnings speaking up in future cockpits.

The author has done experiments on VWS since 1970, both, in flight and in a fighter simulator. The results of these experiments are discussed here, and they show that in spite of some advantages, there are also several human factors problems in the use of VWS that have to be overcome in order to make the second generation VWS as efficient as possible.

RECOGNITION TIME MEASUREMENTS IN HELICOPTER FLIGHT

For this experiment we used a GAF UH-1D helicopter that was equipped with a 20 channel NORVIPS (Nortronics Voice Interruption Priority System) in addition to the standard light warning system (Figure 1). The voice warnings were activated by signals from the standard warning system as well as some extra sensors that had been installed in the helicopter. In addition, all 20 channels of the VWS could be activated by means of push buttons on a simulation box outside the pilot's normal field of vision.

Recognition Time Measurements in Helicopter Flight
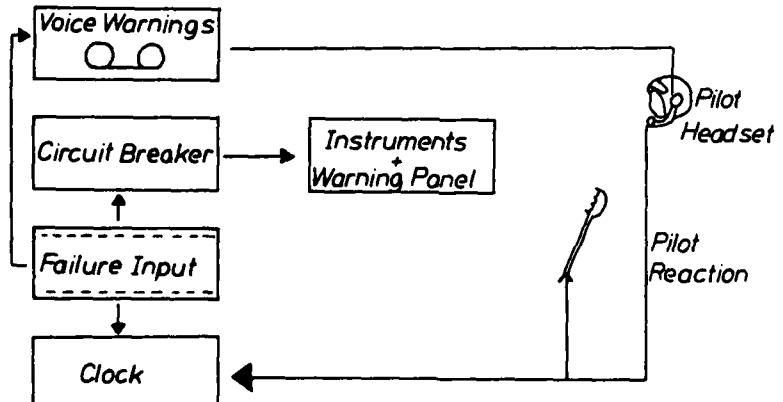
Figure 1

In the experiment the recognition times for light and voice warnings were measured in cruise and nap of the earth (NOE) flights. The simulated failures we used in our experiment are listed in Table I. Whenever possible, the failures were initiated by pulling a circuit breaker on the overhead panel outside the pilot's field of vision. In all other cases the warning was activated by use of the VWS simulation box. The recognition time was measured from the onset of a warning until the pilot's announcement of the recognized failure. Subjects were five pilots of the GAF Testcenter.

Simulated Failures in Helicopter Flight

Light Warnings	Input	Voice Warning Text
Inverter Control	Circuit Breaker	———
Engine Ice Detect	Circuit Breaker + Simulation Box	Check Caution Panel
Fuel Pump Failure	Circuit Breaker	Fuel Pump Failure
———	Simulation Box	EGT high

Table I

LIGHT WARNINGS - RECOGNITION TIME

The recorded recognition times when using light warnings are shown in Table II. On first view, when only comparing the mean times there seems to be no great difference in cruise and NOE flight. This is due to the fact that in two cases (Subject 2 and 4) light warnings were only recognized when an additional voice warning ("check caution panel") was given after 1 minute failure time. In both cases sun glare may have caused the delay. Without these extreme cases the mean recognition time in cruise flight would have been about 4.5 sec. which is well below the mean time of 7.9 sec. in NOE flight.

Light Warnings - Recognition Times (seconds)

Pilot No	1	2	3	4	5	Mean
Flying Time: total/UH-1D	4920/1140	1750/1050	600/300	724/519	1400/80	
Cruise						
Inverter Control	1.0	2.0 2.7	2.1	63.3 38	3.2	11.2
Anti Ice	4.5	1.9 29	1.5	8.0 6.8	2.7	4 4 9.4
Fuel Pump	6.0	5.4 63.3	2.0	3.6 4.7	3.3	12.6
NOE						
Inverter Control	5.5	5.3 8.0	3.3 2.9	5.1 10.9	3.4 2.5	5.4
Anti Ice	9.0	14.8 4.9	6.0 6.6	5.1 14.0	5.5 3.6	7.7 7.9
Fuel Pump	8.3	3.9 7.3	2.3 9.1	35.2 9.7	10.0 10.0	10.6

Table II

VOICE WARNINGS - RECOGNITION TIME

The recorded times (Table III) show a distinct reduction of mean recognition times when using VWS instead of the normal light warnings.

Voice Warnings - Recognition Times (seconds)

Pilot Nº	1	2		3		4		5		Mean	
Flying Time: total/UH-ID	4920/1140	1750/1050		600/300		724/519		1400/80			
Cruise											
Check Caution Panel (Anti Ice)	2.7	2.5		4.2		4.0		4.7		3.6	
Fuel Pump Failure	2.8	2.8		2.4		2.7		3.1		2.8	3.1
EGT high	2.5	2.4		3.0		3.9		3.2		3.0	
NOE											
Check Caution Panel (Anti Ice)	3.6	3.5	2.0	15.2	5.2	7.8	11.7	3.5	11.0	7.0	
Fuel Pump Failure	2.4	2.1	3.0	3.2	2.5	3.3	3.8	2.5	2.8	2.8	4.2
EGT high	2.6	2.2	3.0	2.1	3.5	2.8	3.3	3.2	3.5	2.9	

Table III

Mean times to recognize a voice warning were 3.1 seconds in cruise and 4.2 seconds in NOE flight. As a matter of fact even the difference between recognition times during cruise and NOE flight is deminished, i.e. pilots do recognize failures as fast during NOE flight as in cruise when being alerted by VWS. This is true for all warnings except for the general warning "Check caution panel". The reason is that whenever this warning was activated in the NOE situation the pilots first had to stabilize flight conditions before they could check their instruments and light warning panel.

This finding suggests that such a general warning should be avoided in future VWS.

In addition our data show that it is the less experienced pilot who gains most from the use of voice warnings. Again, this suggests that VWS are of greater value for highly complicated working conditions where operators have to monitor a large number of instruments and displays.

PRIORITY INTERRUPTING SYSTEM

In a questionnaire we also investigated the pilots' preference for different modes of presenting voice warnings. Our pilots suggested a priority interrupting system only for "Fire", "Rotor RPM Low" and "EGT high". For all other failures they recommended having them announced sequentially in the order of their occurence.

Table IV includes a suggested list of failures to be included in a 20 channel VWS for the UH-1D.

Our helicopter-pilots preferred not to have emergency procedures included in the warning text. This is especially remarkable as, in another survey, our jet-fighter pilots said that they could not do without the bold face emergency procedure added to the voice warning.

Suggested Warning List for UH-1D

Priority Interrupting Logic (3 Channels)

1. Fire warning
2. Rotor RPM low
3. EGT high

Time Logic (17 Channels)

4. Transmission chips
5. Engine oil pressure low
6. Transmission oil pressure low
7. Transmission oil hot
8. Hydraulic pressure low
9. Fuel pump failure
10. Engine chips
11. Engine icing
12. 20 minutes fuel remaining
13. Clogged fuel filter
14. DC Generator out
15. Instrument Inverter inoperative
16. Overtorque
17. Governor in emergency position
18. + 19. free for additional warnings
20. Check caution panel
    (aux fuel low, IFF, Engine ice detect out)

Table IV

VOICE WARNINGS AND RADIO COMMUNICATION

During our first investigation we became concerned about the possible interference between voice warnings and radio messages, and we decided to investigate this problem under operational conditions in a F 104 flight simulator (Figure 2). The mission consisted of a navigation flight where pilots had to follow prerecorded ATC radio messages (Table VI). Eleven pilots of a fighter squadron had volunteered as subjects.

Voice Warning Test in F104 - Simulator
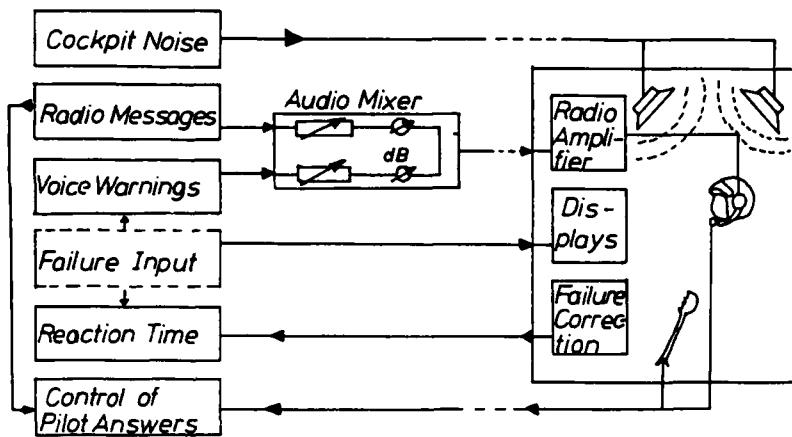
Figure 2

Voice warnings were used alone and combined with light warnings. Again we used a NORVIPS simulation box to introduce the voice warnings.

In this investigation we recorded the actual reaction times. The clock was started at the onset of a failure and the warning, and it was stopped when the pilot initiated the adequate emergency procedure. The failures we used, as well as the warning text and appropriate emergency procedures, are listed in Table V.

In order to investigate the influence of radio communication on reaction times, as well as the effect of voice warnings on the understanding of simultaneous radio messages, part of the failures were introduced while radio communication was going on. Answers and questions by the pilots were used to estimate the amount of correctly understood messages.

The setup of our experiment was so that we could also measure the interphone input level for radio and VWS as selected by the subjects.

Simulated Failures in F104- Simulator

Failures	Voice Warning	Light Warning	Emergency Procedure
EGT high	Check EGT, Check EGT	-	reduce power
Open nozzle	Open nozzle, Open nozzle emergency closure	-	Emergency handle pull
Engine Oil	Engine Oil, Engine Oil	yes	Throttle/Nozzle
APC out	APC out, APC out turn switch off	yes	switch off
Gen № I out	Generator № I out	yes	Attempt reset
Fixed Frequency out	Fixed Frequency out Attempt reset. LN 3 unreliable	yes	Attempt reset

Table V

Radio Messages and Voice Warnings

Radio Message	Failure	Pilot Correct Action
J 22 Ingo taxi runway 25 QNH 1023, report ready		
0 : 30 min		
J 22 is cleared for take-off wind 280°/15 kts, report airborne.		
1 : 10 min		
J 22 climb on (x) heading 250° to FL 80; when reaching FL 80 turn right to heading 360° and continue climb to Fl 250. Report reaching	EGT high (x during message)	250° FL 80 right 360°
3 : 00 min		
J 22 turn left left to heading 270°. Maintain 270° for 1 min. and airspeed 0.9.	Engine oil (while reporting)	270° 0.9. Mach
1 : 00 min		
J 22 a left 360-ty with 25° bank, Throttle 91%; when again on 270°a right (x) 360-ty with 45°bank, airspeed 0.8, contact Monkey Radar on Channel 1	APC out (x during message)	91% 25% left 270° 45° right Channel 1
7 : 10 min		
J 22 right now to 090° and climb to FL 280 squawk 131/343		
3 : 00 min	EGT high (while climbing)	
J 22 left left to heading 100° with 10° bank and descend to FL 190, when reaching FL 190 perform max. climb to FL 400		100° left 10° FL 190 AB FL 400
12 : 20 min	open Nozzle (when reaching FL 400)	
J 22 reduce to 95% and descend with 2000 ft/min to FL 200, when reaching FL 200, turn right to 315° airspeed 0.9, report steady on.		2000/min FL 200 re. 315°
12 : 08 min		
J 22 (x) squawk stand-by	Gen No1 out (x during message)	stand by
0 : 30 min		
J 22 a right 360-ty with 30° bank. When (x) reaching 315° execute a left 360-ty with 20° bank, report steady again	Open Nozzle (x during message)	r. 30° bank 315° 1. 20° bank Report
13 : 00 min		
J 22 turn inbound Ingo TACAN Climb with 1000 ft/min to FL 300		Ingo TACAN 1000/min 300
	Fixed Frequency out (while reporting)	

Table VI

AUDIO LEVEL OF VOICE WARNINGS AND RADIO COMMUNICATION

As each subject used his personal helmet and headset, and as we were only interested in the ratio between audio level of voice warnings and radio communication we did not record the actual audio level at the pilot's ears. The mean input level at the interphone amplifier however was 74.59 dB(A) for radio communication and 71.29 dB(A) for voice warnings. The standard deviation was 1.45 dB(A) for radio and 2.46 dB(A) for voice warnings. These data are especially interesting as all subjects were convinced that they had set the voice warnings at a higher level than their radio. It should be mentioned however that the VWS had a rather good audio quality with a female voice, whereas the radio tape had been recorded with all typical noise and distortion in an actual radio transmission with a male air traffic controller at the microphone.

UNDERSTANDING OF RADIO MESSAGES WITH SIMULTANEOUS VOICE WARNINGS

As we had expected, the pilot's requests to have ATC messages repeated increased considerably when voice warnings were activated simultaneously. Same is true for the number of mistakes done by pilots when fulfilling ATC-orders although our subjects had the option to reduce the number of their mistakes by requesting repeated messages before reacting on ATC-orders.

Under normal conditions with 77 messages to 11 pilots (7 messages to each pilot) only 6 requests for repeat were counted. In the other condition with simultaneous voice warnings the numbers were 44 messages to 11 pilots and 21 requests for repeat.

In order to estimate the number of correctly understood radio messages, we counted pilots' mistakes when fulfilling ATC-orders. In the normal condition only 6 mistakes out of 154 tasks were counted, while 13 mistakes out of 176 tasks were registered in the condition with simultaneous voice warnings.

REACTION TIMES

Pilots' reaction times for correcting failures are listed in Table VII.

Although the differences are not statistically significant, the reaction times suggest following order for different conditions:

voice warning	$\bar{x}$ = 5.2 sec
voice warning + receiving message	$\bar{x}$ = 6.0 sec
voice warning + receiving message + warning light	$\bar{x}$ = 6.5 sec
voice warning + transmitting message + warning light	$\bar{x}$ = 7.6 sec

One remarkable conclusion is, that reaction times apparently can not be reduced when light warnings are added to voice warnings. There may be two possible explanations for our experiment:

a) It might be that the two emergency situations "EGT high" and "open nozzle", we used for voice warnings, had been practiced especially often during simulator flights. In this case the extremely short reaction times could be only a result of excessive training.

According to simulator flight instructors however, this was not the case.

b) Warning lights with simultaneous voice warnings might irritate the pilot so much, that he first wants to do a cross check to verify on the emergency situation before he decides on his reaction. This cognitive process of course would take more time than a mere automatic reaction initiated by a voice warning.

If this is so, our results do not support the rather common assumption that the optimal alert is a combined voice- and light warning.

Voice Warning Reaction Times (seconds) in F 104 - Simulator

Failure	1	2	3	4	5	6	7	8	9	10	11	Mean	
EGT high	4.0	5.0	3.5	4.0	5.0	3.0	3.4	14.0	2.5	12.0	4.0	5.5	5.2
Open Nozzle	4.0	4.0	4.0	5.0	3.0	5.5	4.2	7.0	6.0	6.0	5.0	4.9	
EGT high	10.0	6.5	4.5	7.0	6.5	5.5	3.5	11.0	6.0	5.0	6.0	6.5	6.0
Open Nozzle	3.9	5.0	3.5	4.0	3.0	14.5	4.0	10.0	4.0	4.0	4.0	5.4	
APC out	5.0	5.0	7.5	6.0	15.0	8.1	4.0	11.0	7.0	5.5	6.0	7.3	6.5
Gen No 1 out	3.9	5.5	3.5	5.0	5.0	2.0	5.0	7.0	16.0	6.0	4.0	5.7	
Engine Oil	11.0	13.5	4.0	5.0	8.5	9.0	6.0	5.0	4.0	4.0	6.0	6.9	7.6
Fixed Frequency	4.0	6.0	3.5	4.0	19.0	30.5	4.7	9.0	4.0	4.0	-	8.9	

Header: Pilots No

receive Radio Message — with Light Warning — transmit Radio Message

Table VII

This finding does not contradict other experimenters who have found shortest "reaction times" when using voice warnings together with light warnings. In these experiments "reaction time" was understood to be the time between activation of a warning and pressing of the microphone button by the pilot. Of course, this "reaction time" does not include the mental process to identify a specific failure, to decide on the propper emergency procedure and to initiate it.

A second important finding in our experiment is, that the data prove the disturbing effect of simultaneous radio messages on reaction times. There is no question that this disturbing effect does increase with the mental load of the pilot.

In the condition when our subjects were interrupted by a warning while, transmitting a radio message, the disturbing effect was so strong that in one case a pilot just could not find the propper "reset-button" anymore.

CONCLUSION

In our first experiment, in low level helicopter flight, we have demonstrated the benefits of voice warnings in comparison to light warnings. We could also prove that a precise warning text compared to a general warning does help the pilot considerably to identify a failure.

The second experiment suggests that voice warnings will lose their effectivity when the over all information load of the pilot is very high. Our results show that additional light warnings, for instance, instead of reducing reaction times, tend to slow down pilot's reactions. There is no doubt about the disturbing effect of simultaneous audio signals such as radio messages. Our findings stress the possibility that the pilot might become overloaded when voice warnings do occur while radio communication is going on.

On the bases of these findings the Department of Defense of the Federal Republic of Germany has started a research program in 1978 with the aim to overcome these problems and to enhance the capatibility of future voice warning systems.

REFERENCE

1. USAF "Final Report: Evaluation of Nortronics Voice Interruption Priority System"
   4750 Test Squadron Tyndall AFB, Florida, April 1963

2. BATE, "Cockpit Warning System Comparative Study"
   AMRL-TR-68-193

3. KEMMERLING et al "A Comparison of Voice and Tone Warning Systems as a Funstion of Task Loading"
   Wright Patterson AFB, ASD-TR-69-104

4. LICKLIDER, "Audio Warning Signals for Air Force Weapon Systems"
   WADD-TR-60-814 (1961)

5. REINECKE et al "Untersuchungen mit einem Stimmwarnsystem im Flugsimulator eines Einsatz-flugzeuges" Schriftenreihe FlMedInstLw: Luftfahrtmedizin und Grenzgebiete
   Heft Nr. 31 (1973)

DISCUSSION

DR. R.L. MCKINLEY (UNITED STATES)

Were your subjects actual pilots?

Did the subjects have sufficient training time on voice warning systems?

MR. M.W. REINECKE (GERMANY)

Yes they were. In the UH-ID-experiment we used Test Pilots and in the F104-Simulator experiment we had combat ready fighter pilots.

The pilots ran through a familiarization phase before the experiment. This was when we measured their selected audio level for Voice Warnings and Radio. We did not train the subjects explicitly for fast reactions on Voice Warnings.

The Background and Bases for the Proposed Military
Standard on Acoustical Noise Limits in Helicopters

by
Georges R. Garinther and David C. Hodge
U.S. Army Human Engineering Laboratory
Aberdeen Proving Ground, MD   21005
USA

Summary

A design standard for interior noise of helicopters has been prepared
to provide the developer and user with realistic noise limits which con-
sider hearing damage risk, speech intelligibility, mission profile, state-
of-the-art in noise reduction, and helicopter weight. The levels selected
meet the current hearing conservation limits of the Department of Defense
and permit electrically aided sentence intelligibility of 98%. Helicopters
below 20,000 pounds are treated separately from those above because of the
strong positive relation between internal noise and vehicle gross weight.

This standard defines the locations and flight conditions under which
noise measurements shall be made for compliance. It also specifies the
types of instrumentation and the test procedures to be used to collect in-
terior noise level data. This degree of specificity in the instrumentation
and measurements area is intended to insure that data collected by different
development and test agencies will be both accurate and consistent.

INTRODUCTION

Helicopter noise has been a problem since the helicopter's inception into military
aviation in the late 1940's. As with other types of combat vehicles, excessive noise
in helicopters has a variety of unpleasant or hazardous consequences, including communi-
cation interference and temporary or permanent hearing loss.

Excessive interior noise in helicopters prevents adequate person-to-person and elec-
trically-aided speech communication. Surveys have indicated that in most military heli-
copters the degree of electrically-aided speech intelligibility is below that required
by MIL-STD-1472B (1). This causes problems in command and control, reduces the effective-
ness and response speed of air crews and, in some situations, can actually cause casualties
due to misunderstood instructions.

Person-to-person communication is also often seriously degraded. In troop-carrying
helicopters, a squad leader may only be able to give instructions to his men by shouting.

Noise-induced hearing loss also affects command and control. Temporary losses can
persist long after leaving a vehicle, and this may interfere with the reception of spoken
or whispered commands by troops that have been airlifted to a combat zone. Repeated ex-
posure to excessive noise causes permanent hearing losses, the consequences of which are
well known.

Until now the Army has not had a separate design standard for the noise of helicopters.
Instead, as aircraft, helicopters were supposedly covered by MIL-A-8806A, "Acoustical
Noise Level in Aircraft" (2) despite the fact that MIL-A-8806A was based on fixed-wing
aircraft design, testing mission profiles, power requirements, etc. (Another drawback of
MIL-A-8806A was its use of antiquated criteria for hazardous noise exposure.)

In 1972, when the first version of MIL-STD-1474 (3) was being prepared, the US Army
Surgeon General attempted to include helicopters under the standard (which, of course,
covers all other Army materiel). However, this was met with considerable resistance from
the military aircraft community, and The Surgeon General was persuaded to allow the ex-
clusion of helicopters from the provisions of MIL-STD-1474 on the condition that an effort
be undertaken to develop a separate noise standard for Army helicopters. After several
false starts a US Army Working Group on Helicopter Noise was formed, chaired by Mr. Steve
Moreland of the U.S Army Aviation Research and Development Command. Organizations that
were represented on the Working Group included: The Surgeon General, Training and Doctrine
Command, Human Engineering Laboratory, Environmental Hygiene Agency, Avionics R&D Activity,
Applied Technology Laboratory, and representatives of several major U.S. helicopter manu-
facturers.

CONSIDERATIONS UNDERLYING NOISE LIMITS

The following considerations were addressed during the development of this standard:

1. Potential hearing hazard to both the crew and passengers.

2. Speech intelligibility, both direct and electrically-aided.

3. Attenuation provided by the crew helmet, and hearing protectors which would be worn by passengers.

4. Potential noise exposure to unprotected passengers.

5. State-of-the-art of noise reduction in helicopters.

6. Typical mission profiles of various helicopters in peacetime and in wartime.

7. Weight penalty of adding noise reduction material.

As a preliminary step it was decided to plot the range of octave band levels encountered in various helicopters (4). Figure 1 shows this range along with the current limits of MIL-A-8806A. Also shown for comparative purposes are the octave band levels of a current state-of-the-art helicopter---the Blackhawk. As we shall see later in this paper the limits of MIL-A-8806A were unnecessarily restrictive in the 31.5-125 Hz and the 4000-8000 Hz bands, and too lenient from 250-2000 Hz.
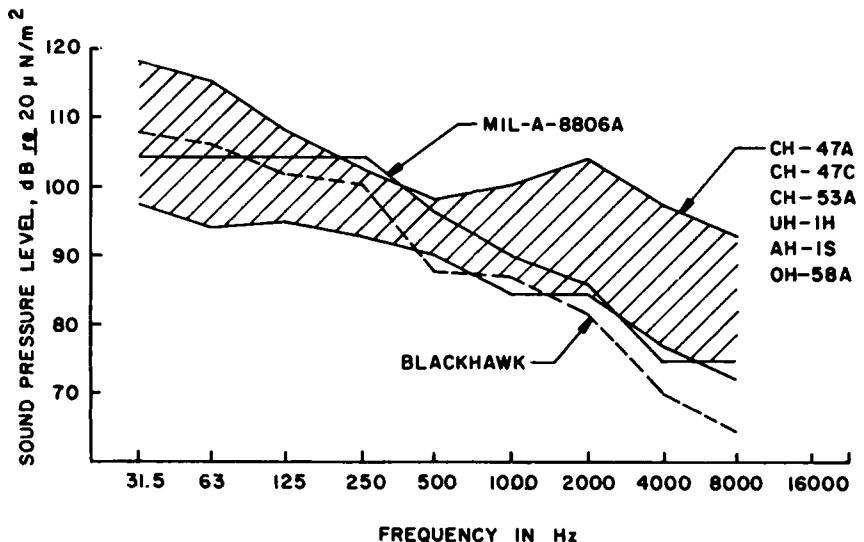
FIGURE 1. Octave Band Pressure Levels in Current Military Helicopters and the Limits of MIL-A-8806A.

It was the intent of the Working Group that the noise limits established would not place undue restrictions on helicopter weight and performance, and that exposure duration during typical missions would be a paramount consideration in determining noise limits. Also, future solutions to noise problems should emphasize noise reduction at the source and limiting acoustic materials in military helicopters to approximately 1% of aircraft gross weight. Emphasis would be placed upon the need to improve helicopter performance while improving communication and hearing conservation.

MISSION PROFILE

Typical noise exposure patterns or scenarios had to be obtained in order to determine the noise levels which should not be exceeded for hearing conservation purposes. Table 1 shows estimates of typical single mission durations for various types of helicopters (5). These data indicate that typical missions range in duration from 1.1 to 2.3 hours.

TABLE 1

Aircrew Exposure Per Mission for Various Helicopters

TYPE HELICOPTER	MISSION PROFILE	MISSION TIME (HOURS)
UTILITY	Combat Troop Extraction	1.3
	Combat Assault	2.3
	Resupply Unit in Contact	1.5
	Aeromedical Evacuation	1.1
SCOUT	Establish Enemy Contact	1.4
	Recon Battle Positions	1.6
	Target Acquisition	1.6
	Screening	2.3
ATTACK	Air-Cav	1.7
	Airmobile Escort	1.6
	Anti-Armor	1.8
CARGO	Logistic Resupply	1.3
ALL	Training	1.8

Table 2 shows the average daily noise exposure times, for a typical number of missions, for peacetime and combat in the four types of helicopters currently in use. Daily aircrew exposure time in combat is two to three times the exposure during peacetime.

TABLE 2

Average Daily Aircrew Exposure during Peacetime
and Combat for Various Helicopters

TYPE HELICOPTER	EXPOSURE TIME (FLYING HOURS)	
	PEACETIME	COMBAT
UTILITY	3-6	8-10
SCOUT	2-4	12-14
ATTACK	2-4	8-10
CARGO	2-4	8-10

Comparing the combat exposure times of Table 2 with current hearing damage risk criteria indicates that the expected 10-12 hours of exposure would require noise levels that are unrealistically low from a helicopter design standpoint. It has also been estimated that 95% of hearing loss in military crews and passengers is incurred during peacetime. For these reasons the maximum daily peacetime crew exposure of 4 hours was selected as the basis for determining the hearing damage limits due to helicopter noise.

Based upon the typical mission times for utility and cargo helicopters shown in Table 1, it was estimated that passengers would be exposed for a period of 1.5 hours per day since they normally only participate in a single mission. It was also estimated that many of these exposures might be without hearing protection. It was concluded, therefore, that the limits of the standard should consider both 4 hour exposures for crew members with hearing protection, and 1.5 hour exposures for passengers without hearing protection.

HEARING CONSERVATION

Department of Defense Instruction 6055.3 (6), upon which U.S. Army hearing conservation programs are based, requires that the 8-hour noise exposure level be below 85 dB(A). The Working Group members agreed that the aircraft noise limits should be specified in terms of octave band levels, however, and Category D from MIL-STD-1474B was established as the baseline limit. The limits of Category D are shown in Table 3; an aircraft noise spectrum which does not exceed these values in any octave band will normally be less than 85 dB(A).

TABLE 3

Damage Risk Criteria for Various Exposures, SPH-4 Attenuation
and Limits of MIL-A-8806A

Octave Band Center Frequency (Hz)	Category D MIL-STD-1474B 8 Hr Unprot. Level (dB)	1.5 Hr Unprot. Level (dB)	4 Hr Prot. Level (SPH-4 minus 1 σ) (dB)	4 Hr Unprot. Level (dB)	SPH-4 Attn. (dB)	MIL-A-8806A (dB)
31.5	117	127	132	121	–	104
63	106	116	122	110	–	104
125	96	106	112	100	17	104
250	89	99	103	93	15	104
500	83	93	112	87	30	96
1K	80	90	108	84	29	90
2K	79	89	123	83	35	86
4K	79	89	129	83	51	75
8K	81	91	129	85	44	75

Department of Defense Instruction 6055.3 also specifies a 4 dB trading relation between time and intensity. Thus, Table 3 also shows the octave band limits for 1.5- and 4-hour exposures, along with the attenuation values for the aircrewman's helmet (SPH-4). For this application the group decided to subtract one standard deviation from these mean attenuation values. Therefore, the octave band limits for a 4-hour daily exposure when wearing the SPH-4 are shown in Table 3. For comparison the limits of MIL-A-8806A are also provided.

Consideration was also given to including the frequencies of 31.5 Hz and 16 kHz in the standard. It was determined that insufficient data and some controversy presently exists for the inclusion of 31.5 Hz. On the other hand it was agreed that the standard should be extended to include 16 kHz and that, based upon research to date, 89 dB for a 4-hour unprotected exposure should be incorporated.

SPEECH INTELLIGIBILITY

Clearly, when establishing a standard for interior noise, an environment which will enable reliable electrically-aided speech intelligibility must be provided. Effective communications is not only important during normal flight, but is particularly critical during flight emergencies when speech intelligibility may radically affect safety of flight.

The sound pressure level selected for hearing conservation purposes provides an articulation index (AI) of 0.6. This AI will yield electrically-aided intelligibility of 84% using the American National Standards Inst., monosyllabic word intelligibility test (7) which corresponds to 98% sentence and standardized phrase intelligibility (8). The achievement of this intelligibility in helicopters is based upon meeting the noise limits of this standard plus using an aircrewman's helmet which provide the electrical and the attenuation characteristics of the current SPH-4 helmet.

Regarding person-to-person speech intelligibility, it was agreed upon by the Working Group that the noise limit of the standard should permit limited communication providing 75% monosyllabic word intelligibility at 1.5 meters when shouting, and at 0.25 meters when using a normal loud voice. This intelligibility is obtainable if the speech intelligibility level (PSIL-4---the mean of 500, 1000, 2000 and 4000 Hz) does not exceed 85 dB.

DESIGN LIMIT FOR HELICOPTERS UNDER 20,000 lbs.

The octave band levels shown in Figure 2 were selected as the design limit which would meet the aforementioned hearing conservation and speech intelligibility requirements. The individual octave band levels were selected as follows:

1. The SPL selected for 63, 125 and 250 Hz are those which meet the hearing conservation limits for unprotected hearing for a 1.5-hour exposure.

2. The SPL selected for 500, 1000, 2000 and 4000 Hz are those which when averaged will meet the 85 dB PSIL-4 requirement for unaided speech intelligibility.

3. The SPL selected for 8000 Hz is that one which will meet the hearing conservation limits for unprotected hearing for a 4-hour exposure.

4. The SPL selected for 16 kHz is that one which will meet the limit adopted by the American Congress of Governmental Industrial Hygienists for a 4-hour unprotected exposure (9).
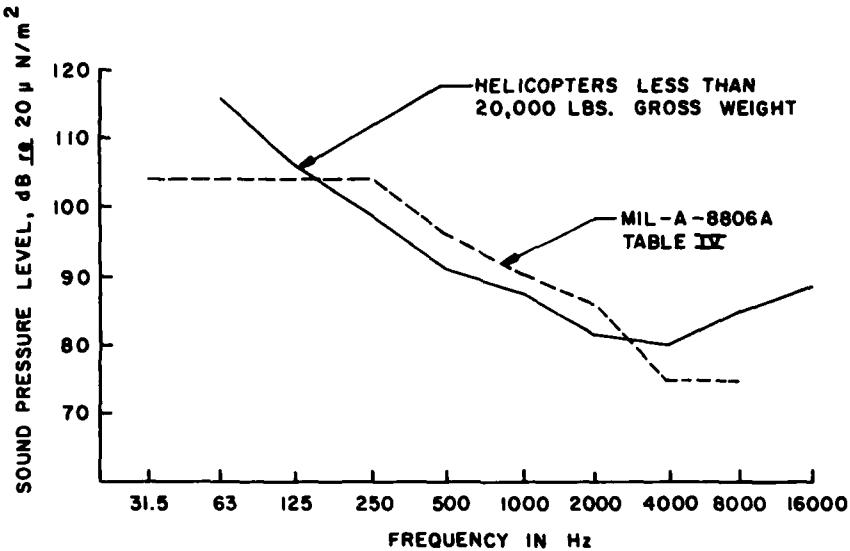
FIGURE 2. Design Limit Helicopters Weighing Less than
20,000 Pounds and the Limit of MIL-A-8806A.

Selection of these levels will also provide adequate electrically-aided speech intel-
ligibility and afford adequate hearing protection for personnel wearing the SPH-4 helmet
for up to eight hours, or approved hearing protectors for up to six hours.

For comparison, Figure 2 also shows the limits of MIL-A-8806A. As can be seen the
low frequency and high frequency limits of that standard were unnecessarily restrictive.
In addition, the levels in the mid-frequencies have been lowered by about 5 dB in order
to improve intelligibility and to decrease the potential for hearing loss.

Examination of the selected design criterion of Figure 2 by the Working Group indi-
cates that helicopters weighing more than 20,000 lbs. could not meet this limit even con-
sidering state-of-the-art advances. Therefore a separate noise limit was derived for
heavy helicopters.

DESIGN LIMIT FOR HELICOPTERS OVER 20,000 lbs.

It was realized when establishing the noise limit for heavy helicopters that a compro-
mise between desirable noise limits and achievable noise levels would be necessary due to
the technical difficulties involved in reducing noise. The approach considered in attempt-
ing to define the noise limit for heavy helicopters was first to estimate the lowest level
achievable through state-of-the-art noise reduction techniques.

As shown in Figure 3, an estimate was made of the reduction in noise level if the
CH-47A (28,000 lbs.) type rear fuselage noise treatment was installed in the CH-47C (40,000
lbs.). These levels were further reduced by the attenuation obtained from a transmission
noise R&D program (10). Assuming an acoustical treatment of 1% of the gross weight, the
estimate is very close to the original quieter CH-47A helicopter data, with elimination
of the peak at 2000 Hz. Plotted through the lower levels of these data is a realistically
achievable level for a heavy helicopter using 1% of gross weight for noise reduction mate-
rials and the latest noise reduction technology.

Superimposed over the realistically achievable level in Figure 4 are the unprotected
hearing conservation limits for 1-½ hours and 45 minutes. Although a 1-½ hour limit would
be a more desirable goal, the achievable levels for this class of helicopter, appear to be
limited to that for the 45 minute exposure. However, this design limit for heavy helicop-
ters would make person-to-person communications at an PSIL-4 of 94 dB almost impossible
with a normal loud voice, and would provide difficult communications even when shouting
at 0.5 meters.

FIGURE 3. Estimated Noise Level Achievable in Heavy Helicopters Using
1% Acoustical Treatment and Using State-of-the-Art Technology.
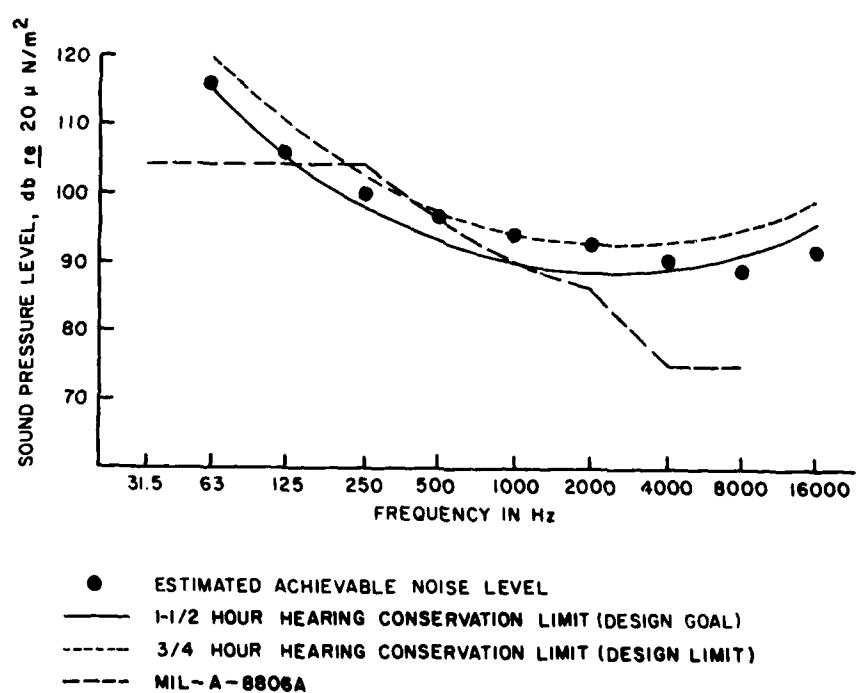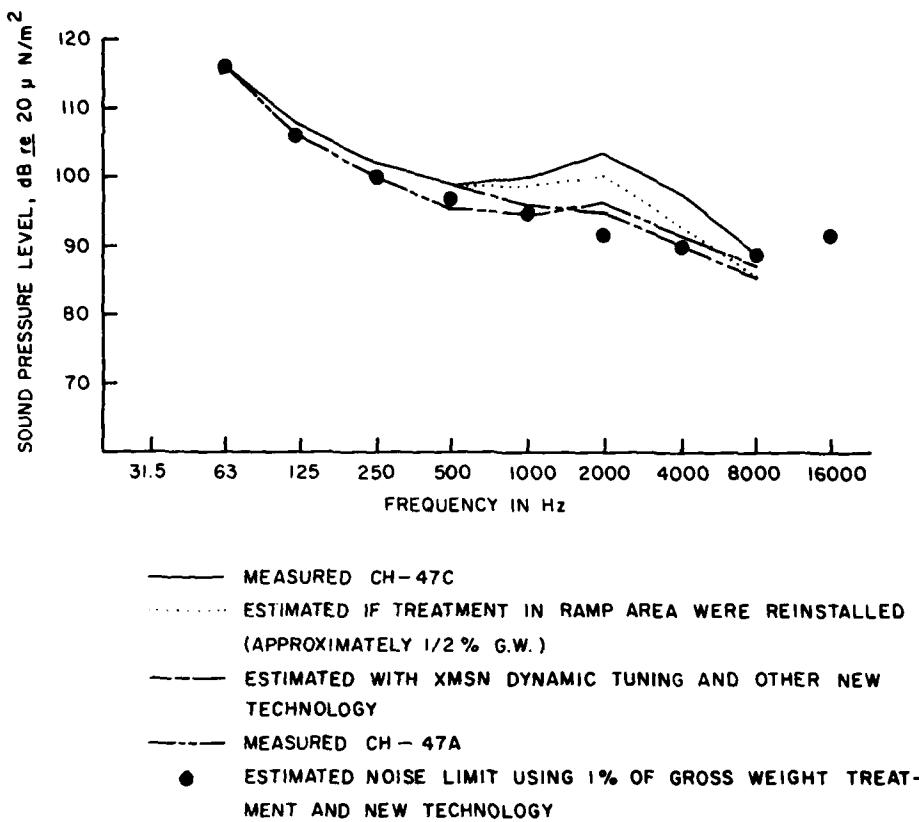
FIGURE 4. Comparison of Hearing Conservation Limit with Estimated
Achievable Noise Level for Heavy Helicopters and the
Limit of MIL-A-8806A.

Comparison of the 1-½ hour and 45 minute limits with levels of very heavy single rotor helicopters such as the CH-53D (35,000 lbs.) and the CH-53E (60,000 lbs.) indicates that they will have difficulty meeting these limits. Figure 5 illustrates the internal noise level actually measured for the CH-53D and projected for the CH-53E (11). It should be noted that in order to meet a design limit the helicopter noise must not exceed the octave band levels at any frequency.
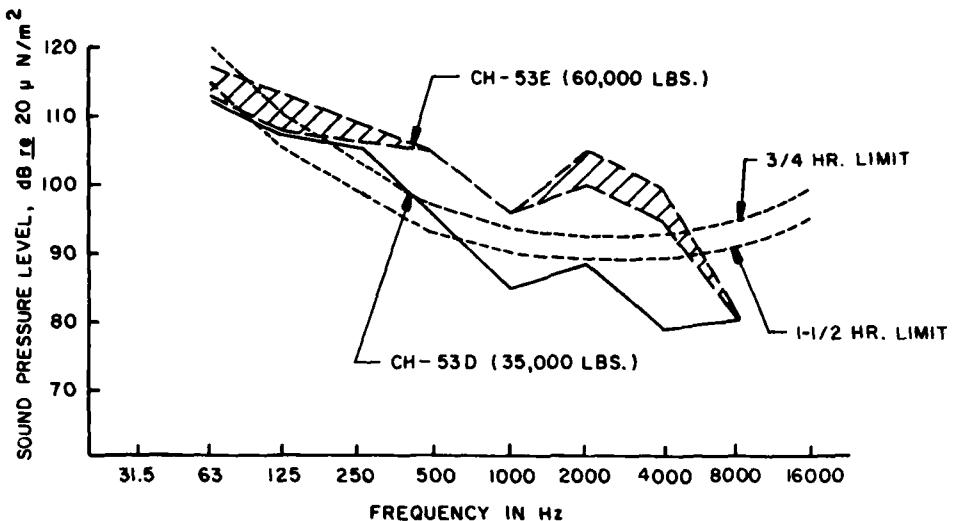

FIGURE 5. Comparison of Hearing Conservation Limit with Noise
Level of Two Single Rotor Heavy Helicopters.

In these large helicopters, the lower frequency range of 63-250 Hz is the most difficult to attenuate because of possible rotor blade performance decrements and large amount of barrier material weight required. The mid-to-high frequencies of 500-8000 Hz can be reduced by direct acoustical treatment of main transmissions during design, and selective use of airframe barrier materials and transmission-to-airframe damping.

For the above reasons and to accommodate the very heavy single-rotor helicopters, two curves were established for helicopters weighing more than 20,000 lbs. The higher is called the "design limit" while the lower curve is the "design goal" for the noise of heavy helicopters, Figure 4 shows these two noise limits as well as the design limit of MIL-A-8806A.

It should be noted that ability to reduce noise for this class of helicopter really drove the design limit curve. The hearing conservation limit selected was determined based on achievable limits rather than using hearing conservation considerations to drive development of the curves. Since the "design limit" curve resulted in a 3/4 hour hearing conservation limit, hearing protection will have to be worn by troops and passengers on any flights over 3/4 hours. For aircrew members wearing the SPH-4 helmet, four hours of daily flight mission exposure are permitted, while for passengers wearing approved hearing protection, three hours of daily exposure are permitted. Table 4 shows the limits by octave band for both light and heavy helicopters.

TABLE 4

Helicopter Sound Pressure Level (dB) Design Limits/Goals

OCTAVE BAND CENTER FREQUENCY (dB)	DESIGN GROSS WT. LESS THAN 20,000 LBS. DESIGN LIMIT	DESIGN GROSS WT. 20,000 LBS. OR GREATER DESIGN GOAL	DESIGN LIMIT
63	116	116	120
125	106	106	110
250	99	99	103
500	91	93	97
1000	87	90	94
2000	82	89	93
4000	80	89	93
8000	85	91	95
16000	89	95	99

The development of the "design goal" was influenced by a desire to meet the 1-½ hour hearing conservation limit required for unprotected passengers as well as to provide the goal of a tolerable environment for unaided communication. It is intended that this "design goal" would be used to advance the state-of-the-art in noise reduction technology and to provide aircraft contractors with a financial or competitive incentive to reduce noise to more acceptable levels.

AIRCRAFT SUBSYSTEM NOISE

In some aircraft subsystem, equipment which operates while in flight such as oil pumps, on-board auxiliary power units, blowers, etc., may produce significant noise levels. Therefore, this standard requires that the noise generated by those systems, while in flight, will not exceed the helicopter noise limit. The fact that government furnished equipment or commercial equipment may significantly contribute to the noise level, does not eliminate the requirement that the total system noise conform to the standard. The contractor is required to apply appropriate noise reduction techniques or place the equipment in such a manner that the noise limits are met.

MEASUREMENT AND EVALUATION

The conditions under which helicopter noise will be measured are separated into two sections in the standard. The first section provides those data that must be obtained to verify contract compliance. For compliance, measurements are made at selected crew and passenger locations when hovering and in normal level flight with all doors, windows and vents closed, and with acoustic treatment in place.

The second set of data are for information purposes and are designed to:

1. Determine the effect of opening doors, windows and vents.

2. Determine the noise exposure of ground and maintenance personnel.

3. Provide data to verify the adequacy of electronic communications equipment.

Noise measurement procedures are precisely defined in the standard, as well as the instrumentation characteristics and the instrumentation techniques to be used. The procedures presented reflect the most current noise measurement techniques available. These techniques were formulated to be restrictive enough to provide consistent, repeatable data from different agencies while providing sufficient latitude to permit the use of various manufacturers' instrumentation. Data to be reported are clearly defined, and include such items as: instrumentation, weather conditions, calibration method, etc.

It is intended that this standard will provide the measurement procedures and noise limits by which reasonable noise levels can be achieved inside helicopters. These noise levels were established to represent a reasonable compromise between those levels which current state-of-the-art permits the designer to achieve in helicopters, and those levels which will minimize hearing damage risk and maximize speech intelligibility.

REFERENCES

1. Department of Defense, "Human Engineering Design Criteria for Military Systems, Equipment and Facilities", MIL-STD-1472B, Wash, DC, 31 Dec 1974.

2. Department of Defense, "General Specification for Acoustical Noise Level in Aircraft", MIL-A-8806A, Wash, DC, 11 July 1966.

3. Department of Defense, "Noise Limits for Army Materiel", MIL-STD-1474B(MI), Wash, DC, 18 June 1979.

4. Camp, Robert, "Data Presented to Working Group on Aircraft Noise". U.S. Army Aeromedical Research Laboratory, Ft. Rucker, AL 36362, 1978.

5. Martin, Ronald S., "Oral Presentation to the Working Group on Aircraft Noise". U.S. Army Aviation Center, Ft. Rucker, AL 36362, 1978.

6. Department of Defense Instruction 6055.3, "Hearing Conservation", Wash, DC, 8 June 1978.

7. American National Standards Institute, "Method for the Measurement of Monosyllabic Word Intelligibility", S3.2-1971, 1430 Broadway, New York, NY 10018, 1971.

8. Kryter, Karl D., "Methods for the Calculation and Use of the Articulation Index", J. Acoust. Soc. Am. 34, 1689-1697, November 1962.

9. Michael, Paul, "An Evaluation of Industrial Acoustic Radiation Above 10,000 Hz", Prepared for the Physical Agents Br. of NIOSH, Feb 1974.

10.   Hartman, R. M., "A Dynamics Approach to Helicopter Transmission Noise Reduction and Improved Reliability". Presented at the 29th Annual Nat. Forum of the Am. Helicopter Soc. (Preprint No. 772), Wash, DC, May 1973.

11.   Schlegel, Ron, "Data Presented to Working Group on Aircraft Noise", Sikorsky Aircraft, Stratford, CT 06602, 1978.

DISCUSSION

MR. P.D. WHEELER (UK)

Table 3 shows an attenuation of 51 dB (mean minus 1) for SPH-4 helmet at 4kHz. This seems rather high. Has the value been corroborated by in-flight measurements?

MR. G.R. GARINTHER (UNITED STATES)

The 4kHz value is high, however it has been verified by the Army's Aeromedical Research Laboratory.

DR. P.A. WILKINS (AUSTRALIA/UK)

Could you please comment on (i) the failure of all of the listed helicopter noise spectra to meet the low frequency requirement of the old standard, (ii) the absence of data at frequencies below the 31.5 Hz octave band in view of the likely upward spread of masking effect from these low frequencies and a consequent reduction in speech intelligibility and audibility of warning sounds.

MR. G.R. GARINTHER (UNITED STATES)

(i) For two reasons, because the old limit was unnecessarily low and because low frequency noise of aircraft is quite high. We hope to reduce this noise in future aircrafts.

(ii) We felt that at this time there is insufficient information to set a limit at 31.5 and below for all three considerations of hearing hazard, intelligibility and warning signals.

THE EFFECT OF NOISE-INDUCED HEARING LOSS ON THE INTELLIGIBILITY OF SPEECH IN NOISE

G.F. Smoorenburg, J.A.P.M. de Laat and R. Plomp

Institute for Perception TNO, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands

SUMMARY

Speech reception thresholds, both in quiet and in noise, and tone audiograms were measured for 14 normal ears (7 subjects) and 44 ears (22 subjects) with noise-induced hearing loss. Maximum hearing loss in the 4-6 kHz region equalled 40 to 90 dB (losses exceeded by 90% and 10%, respectively). Hearing loss for speech in quiet measured with respect to the median speech reception threshold for normal ears ranged from 1.8 dB to 13.4 dB. For speech in noise the numbers are 1.2 dB to 7.0 dB which means that the subjects with noise-induced hearing loss need a 1.2 to 7.0 dB higher signal-to-noise ratio than normals to understand sentences equally well. A hearing loss for speech of 1 dB corresponds to a decrease in sentence intelligibility of 15 to 20%. The relation between hearing handicap conceived as a reduced ability to understand speech and tone audiogram is discussed. It is shown that a hearing loss criterion of 25 dB averaged over the losses at 500, 1000, and 2000 Hz implies a severe handicap. This criterion is often used in the evaluation of hearing impairment. A better criterion is an average value of 15 dB for the losses at 1000, 2000, and 3000 Hz. The higher signal-to-noise ratio needed by people with noise-induced hearing loss to understand speech in noisy environments is shown to be due partly to the decreased bandwidth of their hearing caused by the noise dip.

1. INTRODUCTION

In occupational hygiene hearing is usually checked by measuring tone audiograms. These audiograms give detailed information about changes in sensitivity of the ear to tones of different frequencies. Hearing losses due to excessive noise exposure, for example, show up in a characteristic fashion: they are sharply localized in the high-frequency region. Consequently, they are called "noise dips".

Although the tone audiogram is a powerful tool in characterizing and quantifying hearing loss, it offers no direct information about the hearing handicap. The handicap depends largely on the ability to understand speech both in social situations (at home, at parties, in church) and in professional situations such as aural communication in aviation, the subject of this meeting. In all of these situations speech intelligibility is often impeded by environmental noise. This means that the ability to understand speech is not so much related to the sensitivity of the ear but rather to the ability to discriminate the relevant speech sounds from the noise. The sensitivity of the ear is less important because speech and noise levels are often high in those conditions. People with hearing loss, whether or not using hearing aids, may hear the speech sounds but they may not be able to understand what is being said.

Our research is directed at the ability to understand speech, particularly in noisy situations, and at the relation between speech perception on the one hand and auditory functions such as absolute threshold (tone audiogram), masked threshold for tones (critical ratio) and frequency selectivity of the ear on the other hand. In the present paper we shall focus on the relation between speech perception and tone audiogram with particular reference to hearing loss criteria used in assessing damage risk levels. A future paper will deal more extensively with the relations between speech perception and several auditory functions.

2. TONE AUDIOGRAMS

Auditory thresholds were measured under computer control at 9 frequencies in fixed order: 3000, 4000, 6000, 8000, 12000, 2000, 1000, 500, and 250 Hz. A two-alternative forced-choice paradigm was used to determine thresholds in a reliable way. The test tone was presented during one of two 300-msec intervals indicated by lights and separated by 400 msec. When the interval containing the test tone was indicated incorrectly test-tone level was increased by 3 dB; three successive correct responses were followed by a 3-dB decrease in level. This adaptive procedure converges to a test-tone level at 79% probability of detection.

29 Subjects participated in the experiments, 7 subjects (18-21 years old) with normal hearing and 22 subjects with a noise exposure history. Left and right ears were measured separately. Their data were treated as independent data. Hearing levels of the 14 normal ears were within 5 dB of 0 dB HL measured with conventional audiometers. Fig. 1 shows the results for the 44 ears with noise-induced hearing loss. The hearing losses exceeded by 90, 75, 50, 25, and 10%, respectively, are indicated. The data are given relative to the median hearing level found for the 14 normal ears. Thus, hearing levels were not calibrated physically. In this experimental set up we choose to work with a normal-hearing reference group.
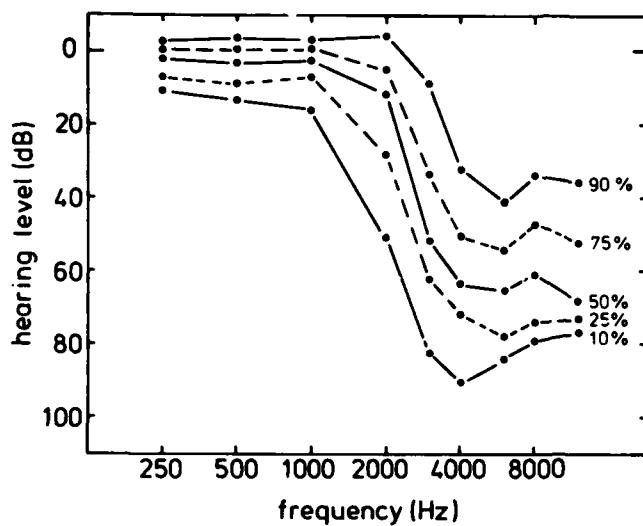
Fig. 1 Hearing loss for 22
subjects (44 ears) with a noise
exposure history exceeded by
90, 75, 50, 25, and 10% of the
ears, respectively. Hearing
loss is expressed relative to
the median threshold found for
14 normal ears.

3. SPEECH RECEPTION THRESHOLDS

Speech reception thresholds (SRT) were measured according to the method developed by Plomp and Mimpen (1, 2). Subjects were presented with simple sentences consisting of eight or nine syllables. Correct reproduction of the entire sentence was a condition for a positive score. An adaptive procedure slightly different from the one described above for the measurement of tone thresholds was used to find the level at which 50% of the sentences were reproduced correctly. The sentences were presented in quiet and in noise at equivalent levels of 25, 40, 55, and 70 dB(A), respectively. The spectral shape of the noise was equal to the average spectrum of the sentences. The method appeared to be very accurate. The standard deviation of the masked speech reception threshold was only about 1 dB. A 1-dB difference in the signal-to-noise ratio corresponds to a change in sentence intelligibility score of 15 to 20%.
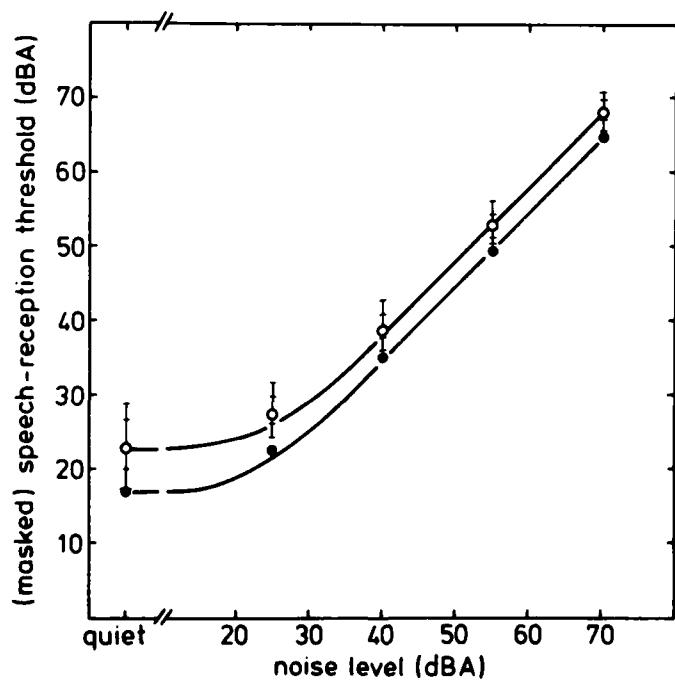
Fig. 2 Median speech reception threshold for normal ears (dots) and for the
ears with noise-induced hearing loss (open circles) as a function of the
noise level and in quiet. The four bars attached to the open circles in-
dicate, from low to high, the speech reception thresholds exceeded by 90,
75, 25, and 10% of the 44 ears with noise-induced hearing loss. The 10-90%
bracket for normals is 7.6 dB in quiet and about 2.8 dB at the higher
noise levels.

The results are presented in Fig. 2. The lower curve describes the data for the 14 normal ears, the upper one the data for the 44 ears with noise-induced hearing loss. At noise levels of 40 dB(A) and higher a nearly constant signal-to-noise ratio is found. The data are fitted by curves according to

$$SRT = 10 \log \{10^{L_O/10} + 10^{(L_N+L_{SN})/10}\} \qquad [1]$$

where L_O is the SRT in quiet, L_N the noise level and L_{SN} the threshold signal-to-noise ratio at high L_N; see Plomp (3). For normal ears we find median values of L_O = 16 dB(A) and L_{SN} = -5.5 dB, whereas Plomp and Mimpen (2) found 19 dB(A) and -5.4 dB, respectively. For noise-induced hearing loss we find median values of L_O = 22.8 dB(A) and L_{SN} = -1.8 dB. Thus, the median signal-to-noise ratio for damaged ears is 3.7 dB higher than the one for normal ears which corresponds to about a 60% decrease in sentence intelligibility score.

4. RELATION BETWEEN TONE AUDIOGRAM AND SPEECH PERCEPTION

When we wish to characterize a tone audiogram with a single number, Fig. 1 shows that the maximum loss found at 4-6 kHz is a good descriptor. The variability of hearing losses in that frequency region is high and the losses at other frequencies are, to some extent, predictable from the maximum loss. Statistical analysis of almost 2000 noise dips (to be reported in a future paper) shows that the best descriptor is a weighted average of the hearing losses with emphasis on the high-frequency losses.

In contrast with the characterization discussed above audiograms are frequently described by thresholds at lower frequencies. For example, a descriptor often used is the average value of the hearing losses at 500, 1000, and 2000 Hz, $HL_{\frac{1}{2},1,2}$. This index is not chosen to get the best characterization of an audiogram but to get some indication of the ability to understand speech. The American Academy of Ophthalmology and Otolaryngology (AAOO) proposed this index as a measure for hearing impairment. According to the AAOO hearing impairment should be evaluated in terms of (dis)ability to hear everyday speech under everyday conditions (4). The ability to hear sentences and repeat them correctly in a quiet environment is taken as satisfactory evidence of correct hearing for everyday speech. Thus, we have arrived at the question of how tone audiogram and speech perception are related to one another.

In the literature (5) the relation between tone audiogram and speech perception is primarily discussed in terms of the significance of hearing loss above 2000 Hz. Whereas the results of some experiments suggest that hearing loss for speech can be adequately predicted from the thresholds at 500, 1000, and 2000 Hz, other results suggest that hearing loss above 2000 Hz is important. Our data show (Table 1) that hearing loss for speech in quiet with the average speech reception threshold for normals as a reference, ΔSRT_O, correlates best with threshold shifts at 500 Hz but that hearing loss for speech in noise averaged over the 40, 55 and 70-dB(A) conditions, $\Delta SRT_{40,55,70}$, correlates best with the shifts in the 2000-3000 Hz region. Since speech intelligibility is often impeded by environmental noise, our finding implies that hearing loss above 2000 Hz is important for speech perception. The difference in correlations found for ΔSRT_O and $\Delta SRT_{40,55,70}$ in Table 1 suggests that the correlation between the two speech shifts themselves is small. Indeed, the correlation coefficient between ΔSRT_O and $\Delta SRT_{40,55,70}$ is only 0.37. This means for noise-induced hearing loss that the speech reception threshold in quiet, often the only speech reception threshold that is measured, gives little indication of the hearing handicap experienced in noisy conditions.

Table 1   Correlation coefficient, ρ, for the hearing loss at several frequencies or combination of frequencies and the hearing loss for speech in quiet, ΔSRT_O, or in 40, 55, and 70 dB(A) noise, $\Delta SRT_{40,55,70}$.

frequency in Hz	ΔSRT_O ρ	$\Delta SRT_{40,55,70}$ ρ
500	.63	.07
1000	.46	.34
2000	.26	.71
3000	.31	.67
4000	.17	.53
6000	.21	.42
500,1000,2000	.52	.68
1000,2000,3000	.36	.74

As mentioned before the AAOO proposed the index $HL_{\frac{1}{2},1,2}$ to quantify hearing impairment where hearing impairment was defined in terms of (dis)ability to repeat sentences correctly in a quiet environment. The AAOO proposed a "fence" of 25 dB $HL_{\frac{1}{2},1,2}$ above which hearing is considered impaired. This fence was adopted by ISO/R-1999 (6). The relation between $HL_{\frac{1}{2},1,2}$ and hearing loss for speech in noise, $\Delta SRT_{40,55,70}$, is given in Fig. 3, the correlation coefficient is 0.68 (Table 1). The results show that a fence of 25 dB $HL_{\frac{1}{2},1,2}$ corresponds to values of $\Delta SRT_{40,55,70}$ above 6 dB. This is a dramatic increase in speech reception threshold. At certain signal-to-noise ratios it corresponds to a decrease in sentence score from about 100% down to 0%. Technically, an increase of signal-to-noise ratio by more than 6 dB required to compensate for the hearing loss is often difficult to achieve. Our data for noise-induced hearing loss suggest that 25 dB $HL_{\frac{1}{2},1,2}$ is clearly too high a fence for hearing impairment.
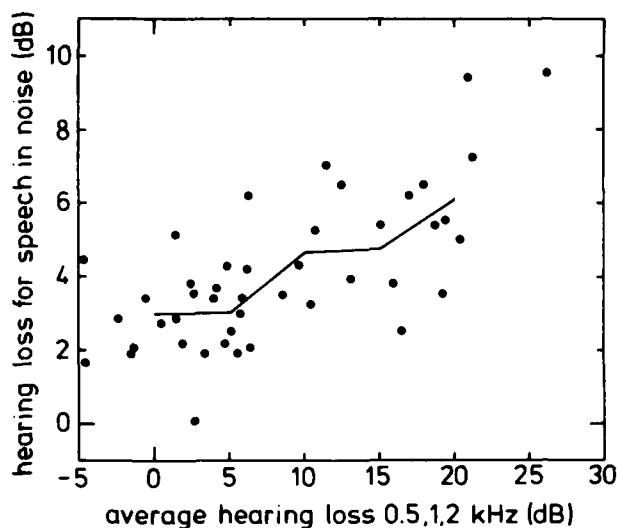
Fig. 3 Average hearing loss for speech in noise at levels of 40, 55, and 70 dB(A) as a function of the average value of the hearing losses in the tone audiogram at 500, 1000, and 2000 Hz for 44 ears. The curve connects average values for hearing loss classes 5 dB in width.

Kryter (5) has proposed new measures of hearing impairment based on the average value of the hearing losses at 1000, 2000, and 3000 Hz, $HL_{1,2,3}$. He proposed fences at 10, 15, and 25 dB for weak, conversational and everyday unamplified speech in quiet. Smoorenburg (7), in deriving damage risk criteria for impulse noise, set the fence at 15 dB $HL_{1,2,3}$. The relation between $HL_{1,2,3}$ and hearing loss for speech in noise ($\Delta SRT_{40,55,70}$) is given in Fig. 4, the correlation coefficient is 0.74 (Table 1). At 15 dB $HL_{1,2,3}$ Fig. 3 shows values of $\Delta SRT_{40,55,70}$ between 2 and 4 dB. In our opinion a higher value than 3 dB cannot be tolerated from a hearing conservation point of view. $\Delta SRT_{40,55,70}$ = 3 dB means a critical hearing distance (50% sentence intelligibility) of 70 cm if it is 100 cm for normals in the same situation. Moreover, for aging normals 3 dB median $\Delta SRT_{40,55,70}$ is not found until about 65 years (2). An upper limit of 3 dB for $\Delta SRT_{40,55,70}$ implies that the fence should not exceed 15 dB $HL_{1,2,3}$.

Fröhlich (8) mentions that hearing losses for pilots in the German Armed Forces should not exceed 30 dB in the frequency region 250-2000 Hz. In this case the criterion is not set for hearing conservation purposes but for professional performance. In our data a maximum of 30 dB HL over 250-2000 Hz corresponds to 16 dB $HL_{\frac{1}{2},1,2}$ or 35 dB $HL_{1,2,3}$ (about the 25% curve in Fig. 1). The limit means a hearing loss for speech in noise of 3 to 7 dB. This loss implies a hearing handicap. However, the professional handicap with respect to pilots with normal hearing will be smaller than 3 to 7 dB because the bandwidth of communication channels is often limited to 3000 Hz.

A misleading comment on speech perception in noise for people with noise-induced hearing loss is made by Kryter (5). He suggests that people suffer no handicap from their hearing impairment when they are in the noise that caused the hearing loss. Kryter reasons that noise-induced hearing loss is usually smaller than the level of the noise inducing

Fig. 4 Average hearing loss for speech in noise at levels of 40, 55, and 70 dB(A) as a function of the average value of the hearing losses in the tone audiogram at 1000, 2000, and 3000 Hz for 44 ears. Note the different scale along the abscissa compared to Fig. 3. The curve connects average values for hearing loss classes 10 dB in width.
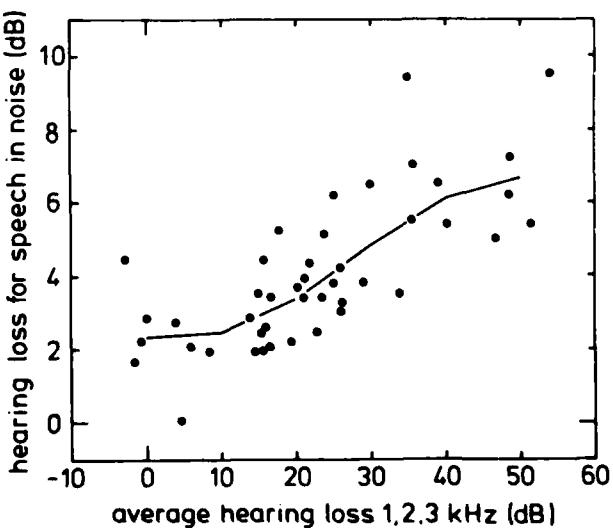
that hearing loss. Thus, sounds will still be heard. Above, we have shown that higher signal-to-noise ratios are required for people with noise-induced hearing loss than for those without hearing loss in order to score 50% sentence intelligibility. Consequently, people with noise-induced hearing loss are handicapped in perceiving speech, or e.g. emergency signals, while they work in the noise causing the loss.

In this section we have discussed the relation between the tone audiogram and speech reception threshold in noise. The correlation between the two is limited. This means that predictions of the speech reception threshold from the tone audiogram have only limited accuracy. Determination of the hearing handicap should therefore imply measurement of speech reception thresholds, including those in noise.

5. ORIGIN OF HEARING LOSS FOR SPEECH IN NOISE

In the preceding section the highest (positive) correlation between hearing loss for speech in noise and hearing losses in the tone audiogram was found in the 2000-3000 Hz region of the audiogram. This suggests that the position of the slope of the noise dips is important. A large frequency region with little hearing loss correlates with small hearing loss for speech. For people with normal hearing we know that bandwidth reduction impedes speech intelligibility. The negative effect on speech intelligibility can be counteracted by improving the signal-to-noise ratio in the channel with reduced bandwidth. It seems that this relation between bandwidth and signal-to-noise ratio is also present in our data for noise-induced hearing loss. In this section we shall examine to what extent the relation between tone audiogram and hearing loss for speech can be understood on the basis of the assumption that hearing loss impedes speech intelligibility in the same way as bandwidth reduction does for normals. For this purpose we shall use a method to calculate speech intelligibility given by Kryter (9) after French and Steinberg which involves calculation of the so-called Articulation Index (AI).

The AI is defined as

$$AI = \sum_{i=1}^{15} \alpha_i (L_{SN,i}+12), \quad \sum_{i=1}^{15} \alpha_i = 0.0333 \qquad [2]$$

where α_i is a weighting factor, $L_{SN,i}$ is the signal-to-noise ratio in dB and i indicates one of fifteen 1/3-octave bands from 200 Hz up to 5000 Hz. The signal-to-noise ratio in each 1/3-octave band equals the overall signal-to-noise ratio because the shape of the noise spectrum equals that of the speech, hence

$$AI = (L_{SN}+12) \cdot \sum_{i=1}^{15} \alpha_i \qquad [3]$$

At 50% sentence intelligibility $L_{SN} = SRT - L_N$. The results for normals having the full bandwidth at their disposal ($\Sigma \alpha_i = 0.0333$) are given in Table 2.

Table 2 Articulation Index AI and its standard deviation σAI for 14 normal ears calculated from the measured signal-to-noise ratio L_{SN}.

L_n dB(A)	L_{SN} dB	σL_{SN} dB	AI	σAI
40	-5.00	1.31	0.233	0.044
55	-5.91	0.92	0.203	0.031
70	-5.47	1.12	0.218	0.037
average of 40, 55, 70	-5.46	1.13	0.218	0.038

If we assume that the tone threshold is caused by internal noise of the ear the AI can also be calculated for the speech reception threshold in quiet from $L_{SN,i} = SRT_i - L_{N,i}$ where SRT_i is the speech level at 50% sentence intelligibility in each 1/3-octave band and $L_{N,i}$ is the noise level per 1/3-octave band that would produce the audiometric threshold. Since speech spectrum and audiometric threshold have different shapes we have to calculate L_{SN} per 1/3-octave band for this case. We find the same AI (0.218) as for the speech reception thresholds in noise. Thus, the proposed link between audiometric threshold and masking noise level leads to consistent AI values for SRT_0 and $SRT_{40,55,70}$. We shall follow this line to calculate the AI for noise-induced hearing loss.

Fig. 5 shows the hearing levels with the median threshold for normals as the reference level presented before in Fig. 1, this time plotted as increased thresholds. Also plotted are the noise spectra at 40, 55, and 70 dB(A), respectively, and the spectrum at the median speech reception threshold in quiet for the 44 ears with noise-induced hearing loss ($SRT_0 = 22.8$ dB(A)). The spectral levels given in Fig. 5 are the masked thresholds for the noise. Thus, all curves in Fig. 5 represent tone thresholds. The AI was calculated for each ear individually on the basis of the measured SRT and the frequency at which the noise spectrum intersects the audiogram. Above this frequency the audiogram increases rapidly which implies a rapid decrease of the apparent signal-to-noise ratio and thus rapidly vanishing contributions to the AI from this frequency region. Table 3 shows the averaged results.

At the three noise levels, L_N, the AI found for noise-induced hearing loss appears to be higher than the AI found for normal ears (0.218). (The standard deviation of the average AI is only $0.040/\sqrt{43} = 0.006$.) Also, the AI calculated for the median speech reception threshold in quiet found for noise-induced hearing loss is higher. Its value is
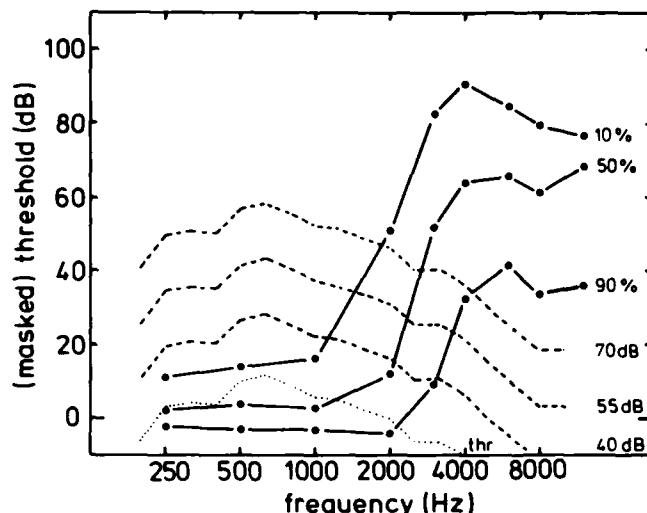
Fig. 5 Audiograms as in Fig. 1 plotted as increased thresholds together with the spectra of the 40, 55, and 70 dB(A) noises used in the speech reception threshold measurements. The spectral noise levels indicated correspond to the threshold levels of tones in that noise. Also indicated is the spectral level corresponding to the speech reception threshold in quiet (thr.). Speech and noise spectra had equal shapes. 0 dB corresponds to the median tone threshold for normal ears.

Table 3 Average values of the measured signal-to-noise ratio L_{SN} and its standard deviation σL_{SN}, the cut-off frequency f of the effective bandwidth for ears with noise-induced hearing loss and its standard deviation σf, the calculated articulation index AI and its standard deviation σAI and the predicted signal-to-noise ratio L'_{SN} assuming that the reduced bandwidth for noise-induced hearing loss is compensated completely by an increased signal-to-noise ratio.

L_n dB(A)	L_{SN} dB	σL_{SN} dB	f Hz	σf Hz	AI	σAI	L'_{SN} dB
40	-0.68	2.75	2010	570	0.233	0.041	-1.79
55	-1.87	2.23	2420	620	0.241	0.044	-3.03
70	-1.62	2.06	2860	760	0.270	0.040	-3.81

0.242 which is close to the values found for L_N = 40 and 55 dB(A). These results imply that hearing loss for speech, ΔSRT, cannot be explained completely by bandwidth reduction as revealed in the audiogram. Yet, bandwidth reduction does explain part of ΔSRT. For noise-induced hearing loss we find an increase in L_{SN} of 4.32, 4.04, and 3.85 dB for L_N = 40, 55, and 70 dB(A), respectively. The increase predicted from bandwidth reduction is 3.67, 2.43, and 1.65 dB, respectively (compare L'_{SN} in Table 3 with the average value of L_{SN} for L_N = 40, 55, and 70 dB(A) in Table 2). Thus, we have not accounted for differences of 0.65, 1.61, and 2.20 dB, respectively. The increasing difference suggests that the effective bandwidth does increase less with increase L_N than what may be expected from the audiogram. Bandwidth reduction also explains part of the variance in the data. For L_N = 40, 55, and 70 dB(A) we find σL_{SN} = 2.75, 2.23, and 2.06 dB, respectively. Disregarding bandwidth reduction σAI would be 0.092, 0.074, and 0.069, respectively, whereas we found 0.041, 0.044, and 0.040. The ratio of the variances, F, are 4.9, 2.9, and 2.9, respectively. With 43 degrees of freedom this means that σAI found when bandwidth reduction is included is significantly smaller than when it is not. The probability that the variances with and without including bandwidth reduction are equal is smaller than 1%.

In describing his SRT data Plomp (2, 3) distinguishes between two types of hearing loss for speech: The "attenuation" loss SHL_A and the "distortion" loss SHL_D. His model implies a constant ΔSRT at the higher noise levels which then equals SHL_D. Our data fit his model well (Fig. 2 and Table 3). In this section we have shown that part of SHL_D is probably due to bandwidth reduction. The increase in SHL_D not accounted for is subject of further investigations.

REFERENCES

(1) Plomp, R., and Mimpen, A.M., Improving the reliability of testing the speech reception threshold for sentences, Audiology 18, 1979, 43-52.

(2) Plomp, R., and Mimpen, A.M., Speech-reception threshold for sentences as a function of age and noise level, J. Acoust. Soc. Am. 66, 1979, 1333-1342.

(3) Plomp, R., Auditory handicap of hearing impairment and the limited benefit of hearing aids, J. Acoust. Soc. Am. 63, 1978, 533-549.

(4) Glorig, A., and Baugh, W.L., Basis for percent risk table, Proc. Int. Congr. on Noise as a Public Health Problem, Superintendent of Documents, U.S. Government Printing Of-

fice, Washington D.C. 20402, 79-102.

(5) Kryter, K.D., Effects of Noise on Man, Ch. 4, 1970, Academic Press Inc., New York.

(6) International Organization for Standardization, Recommendation 1999, Assessment of occupational noise exposure for hearing conservation purposes, 1971.

(7) Smoorenburg, G.F., Damage risk criteria for impulse noise, Proc. Int. Symp. on New Perspectives on Noise Induced Hearing Loss, Cazenovia, New York, June 1980, to be published by Raven Press, New York, 1981.

(8) Fröhlich, G., Ton- und Sprachgehör des alternden Flugzeugführers, Arbeitsmedizin-Socialmedizin-Arbeitshygiene 9, 1972, 261-266.

(9) Kryter, K.D., Effects of Noise on Man, Ch. 2, 1970, Academic Press Inc., New York.

DISCUSSION

DR. G.M. ROOD (UK)

Could I ask how TNO calculates the signal-to-noise ratios when using speech or considering the impulsive nature of speech?

DR. G.P. SMOORENBURG (NETHERLANDS)

We calculate the energy contained in the speech by numerical integration of the speech signal sampled at a high rate. (The actual waveform is sampled. There is no integrating detector circuit before sampling.)

DR. P.A. WILKINS (AUSTRALIA/UK)

To what extent do you think that a similar test of speech intelligibility in noise with binaural (dichotic) listening in a semi-reverberant environment would

(i) improve the sensitivity of the measure?
(ii) indicate a greater hearing impairment with noise-induced hearing loss?

DR. G.P. SMOORENBURG (NETHERLANDS)

In our experience binaural listening doesn't improve the sensitivity of the measure but rather we find a lower signal-to-noise ratio at 50% sentence intelligibility. In other words, the curve giving percentage of sentences correctly understood as a function of signal-to-noise ratio doesn't become steeper but it shifts to lower signal-to-noise ratios.

If you mean by improved sensitivity of the measure that additional aspects such as binaural integration are measured too with binaural listening, I agree. For ambient noise conditions binaural integration may result in about 2 dB lower signal-to-noise ratios. In case of one competing speaker larger effects may be expected depending on the angle between the speakers.

DR. R.L. MCKINLEY (UNITED STATES)

Could you briefly describe articulation index.

DR. G.P. SMOORENBURG (NETHERLANDS)

The articulation index is an index which calculates the speech-intelligibility on the basis of the signal-to-noise ratio in a number of third octaves. The signal-to-noise ratio may be in between -12 dB which means no contribution to speech intelligibility, up to +18 dB which means full contribution for that third octave. Then every third octave is weighted according to the curve shown in this slide. Then if you sum the signal-to-noise ratio in each third octave, with the appropriate weighting you get the articulation index and then there is a relation between the articulation index on the X-axis and the speech intelligibility on the Y-axis.

HEARING STANDARDS FOR AIRCREW

by

Martin P.C.Gloudemans
Captain RNLAF
Bio-Environmental Engineer/Industrial Hygienist at the Aeromedical
Department of the Royal Netherlands Airforce
Kampweg 2, Postbus 153, 3769 ZK Soesterberg
Netherlands

Almost a year ago, I read a call for papers, stating: "In modern military aircraft, it is mandatory that aircrew should be able to perceive and respond to audio information, whether this be speech or tones, with minimum effort and highest reliability."

Well, it goes without saying, that, in order to comply with that statement, it is not only necessary to have high quality airborne voice communication systems, but also crewmembers with a very good hearing ability.
While preparing this paper I was confronted with a number of national hearing standards for aircrew, which happen to show important differences.

Table 1 shows the Dutch hearing standards for aircrew:

Authority	Maximum allowable hearing loss (dB, re. ISO-1964) in the frequencies:						
	250	500	1000	2000	3000	4000	5000
Rijksluchtvaartdienst (= Netherlands Civil Aviation Authority)	--	35	35	35	50	--	--
Rijks Geneeskundige Dienst (= Authority in charge of medical examination of the Dutch civil servants)							
a. initial examination	--	10	10	10	30	--	--
b. periodic examination	--	30	30	30	45	--	--
Koninklijke Marine en Koninklijke Luchtmacht (Royal Netherlands Navy and Royal Netherlands Airforce)							
a. initial examination	20	20	20	20	30	30	30
	in case of assymetry: mean loss of 20 dB and a maximum of 30 dB for the poorer ear.						
b. periodic examination	20	20	20	20	20	--	--

TABLE 1, Hearing standards for aircrew in the Netherlands.

It is remarkable that in the Dutch Armed Forces on periodic examinations, the maximum allowable hearing loss at the 3000 Hz-frequency, is less than on initial examination! One will not be surprised to learn that many waivers are granted and that hearing standards are in review nowadays.

In addition to the figures in table 1, some other requirements are formulated:
The Rijksluchtvaartdienst, declares an applicant with a hearing loss greater than indicated in table 1, fit to fly, provided that:

a. the applicant has a hearing performance in each ear separately equivalent to that of a normal person, against a background noise that will simulate the masking properties of flight deck noise upon speech and beacon signals;

  and:

b. the applicant has an ability to hear an average conversational voice in a quiet room, using both ears, at a distance of 2 metres from the examiner, with the back turned to the examiner.

The Rijks Geneeskundige Dienst, makes an exception at initial examination for the cases wherein the origin of the hearing loss precludes a further progression of the hearing loss. In these cases the requirements for the periodic examination apply.

The Netherlands Armed Forces, have for the periodic examinations, the additional requirement of speech intelligibility. The maximum acceptable loss is 20 dB in a masking noise of 90 dB.

Observing so little unity within a small country as the Netherlands, it may be revealing to have a look at the various hearing standards in other Nato countries.

Table 2 is a summary of -on request obtained- audiometric hearing standards in Italy, Portugal, Norway, The United States of America, Canada and France.

Country	Maximum allowable hearing loss (dB, re. ISO-1964) in the frequencies (Hz)									
	125	250	500	1000	2000	3000	4000	5000	6000	8000
Italy	10	10	10	10	10	10	10	10	10	10
Portugal, both ears	--	15	15	15	15	--	:	--	--	:
							average 40			
Canada, each ear	--	--	30	30	30	30	--	--	--	--
Norway and USAF										
a. Flying classes I and IA, each ear	--	--	25	25	25	:	:	--	:	--
							average 45			
b. Flying classes II and III,										
– better ear	--	--	30	30	30	--	--	--	--	--
– poorer ear	--	--	30	50	50	--	--	--	--	--
US Army										
a. classes 1 and 1A										
– each ear	--	--	25	25	25	55	55	55	--	--
b. class 2										
– better ear	--	--	25	25	25	35	35	35	--	--
– poorer ear	--	--	25	35	35	45	45	45	--	--
c. class 3										
– better ear	--	--	30	25	25	35	35	35	--	--
– poorer ear	---------------no requirements---------------									
US Navy, each ear	--	--	25	25	25	45	60	--	--	--
France, each ear										
a. hearing standard 1:	--	35	35	30	30	30	35	--	--	40
b. hearing standard 2:	--	45	45	40	40	50	55	--	--	--
c. hearing standard 3:	--	55	55	50	50	60	65	--	--	--

<u>TABLE 2</u>, Hearing standard for aircrew in six NATO countries.

The French hearing standards have not altered from those published in AGARDograph No 213. The figures in table 2 are the -to ISO-1964- decibels converted, standards from that document.

Once again there are additional requirements.
Italy states:

a. whispered voice must be heard at 9-7-5- metres respectively for high, medium and low frequencies in each ear separately;
b. in alternative, a regular watch must be heard in each ear separately at one metre;
c. tested on a pure tone audiometer against a masking noise, the candidate shall not have a hearing loss greater than 20 dB at the frequencies indicated in table 2;
d. the candidate shall have a stereophonic perception at least 3 metres on equidistant sources of sound;
e. for experienced aircrew, flexible criteria of judgement are applied and waivers are granted.

Norway has adopted the US Hearing Standard for Aircrew. Aircrew whose hearing loss approach the maximum loss, tolerated for continued service, are subjected to tests for speech comprehension in noise, applied in a relevant flying situation; which means that a test is done if they respond appropriately to flight orders from the ground, transmitted by radio through their headset during actual flight.

France.

Hearing Standard 1: The limit may be increased to 50 dB at the frequency 3000 Hz, and to 55 dB at the frequency 4000 Hz in the case of regular military personnel who apply for employment as flying personnel. All such applicants undergo a speech intelligibility test which must meet the requirements for Hearing Standard 2.

Hearing Stanard 2: If the audiometric test does not give satisfactory results, a speech intelligibility test is conducted with and without masking noise, either in a free field or with headphones, both ears being tested simultaneously. The sound level in the chamber is approximately 85 dB for the free field test, and 65 dB for the test with headphones. The characteristics of the curves obtained are defined as follows:
a curve, the slope of which is sufficient to achieve 100% intelligibility in 30 dB, with a loss at the 50% threshold, not exceeding:
a. 25 dB, in the test without masking noise, and:
b. 15 dB, in the test with masking noise.

Hearing Standard 3: If the audiometric test does not give satisfactory results, a speech intelligibility test is performed under the same conditions as for Hearing Standard 2.
The characteristics of the curves obtained are defined as follows:
a curve, the slope of which is sufficient to achieve 100% intelligibility in 40 dB, with a loss at the 50% threshold, not exceeding:
a. 30 dB, in the test without masking noise, and:
b. 20 dB, in the test with masking noise.

Only the United States Airforce and the United States Army explicity state that the audiometers should be calibrated in accordance with the ISO-1964 standard.
One may assume, however, that all European audiometers are also calibrated to that ISO-1964 standard. Anyhow, the Dutch audiometers are.

The French figures from table 2 seem rather high. This may be due to conversion from ASA-decibels to ISO-decibels, by adding the following values:

Frequency (Hz)	Values to be added, in order to convert ADA-dB to ISO-dB (From AGARDograph no 213)
250	15
500	15
1000	10
2000	10
3000	10
4000	5
6000	10
8000	10

TABLE 3, Conversion ASA-dB to ISO-dB.

Another interesting fact, noticed when studying hearing standards, was that the, since AGARDograph No 213 changed, United States requirements are apparantly less stringent than those published in that particular document. However, due to the audiometric zero-calibration, the requirements appear to be less stringent only for the United States Navy and then only for the frequencies 500 and 4000 Hz. For the United States Airforce and the United States Army -and also Canada- ,however, the requirements are for the 500 Hz-frequency in fact 5 dB more stringent, and unchanged for the other frequencies.
I wonder why the United States Navy decides to become more tolerant for individuals with hearing loss at 500 Hz, while the United States Airforce and Army change their standard in a less tolerant one.

This practice, and the demonstrated differences, not only between the Nato-countries, but even within the boundaries of one and the same country, leads me to the conclusion that hearing standards for aircrew should be reviewed and standardized. I think that flight-surgeons who declare pilots and other aircrew fit to fly, in spite of the inability to meet the audiometric requirements, often grant the waivers on a very subjective base.
There are, however, tests for speech comprehension in noise, which can be used as a supplement to audiometry. The results of such tests are -provided they are standardized- in my opinion a good criterion when judging an individual's ability to perform his task as an aircrew member.
At a next lecture in this meeting -that will be on Thursday-morning- examples of such tests will be presented by the representatives of the Norwegian Institute of Aviation Medicine.
I would like to pledge for a worldwide introduction of such a test, the results of which should be the basis for further admission of experienced but deafened aircrew.

DISCUSSION

COL. J. CLEMENT (BELGIUM)

Should there still exist any standard for flying fitness based on tonal audiometry in members of flying personnel? Is speech comprehension not sufficient? This question does not pertain of course to applicants.

CAPT. M.P.C. GLOUDEMANS (NETHERLANDS)

In my opinion there is no need for pure tone audiometry in assessing the hearing performance of flying personnel. Pure tone audiometry is essential for hearing conservation purposes, however, and should not be removed from the list of initial and periodic examinations for aircrew.

To judge the ability to perform a task as an aircrew-member, a standardized speech comprehension test in a (simulated) cockpit-noise-environment is preferable to tonal audiometry.

LTCOL J.L. GOLDSTEIN (UNITED STATES)

There was a change in U.S. Army standards which corrected the error in the Class II levels. Pure tone audiometry monitors for hearing conservation purposes and provides a guide to the flight surgeon as to when to request additional evaluation of the aviator.

CAPT. M.P.C. GLOUDEMANS (NETHERLANDS)

Thank you for this additional information.

Of course, tone-audiometry is useful for hearing conservation purposes. I don't think, however, that we need pure-tone audiometry in order to get information about one's hearing ability in connection with his or her occupational task in flight.

COMPARATIVE INTELLIGIBILITY OF SPEECH MATERIALS PROCESSED BY STANDARD AIR FORCE VOICE COMMUNICATION SYSTEMS IN THE PRESENCE OF SIMULATED COCKPIT NOISE

Thomas J. Moore, Ph.D.
Charles W. Nixon, Ph.D.
Richard L. McKinley, B.S.
Air Force Aerospace Medical Research Laboratory
Wright-Patterson Air Force Base, Ohio 45433

SUMMARY

This paper describes the effects of different levels of a simulated operational noise environment on the intelligibility of standardized speech materials as processed through representative US Air Force voice communication systems. Among the systems evaluated was the ARC-164 radio which will serve as the reference system against which the performance of jam-resistant, secure systems developed in the immediate future will be compared. This paper will also report on relative differences found between male and female talkers under various levels of simulated cockpit noise.

INTRODUCTION

A major requirement for the evaluation of a voice communication system is the ability to make a reasonable prediction of the intelligibility of speech signals processed through the system. To achieve this estimate empirical intelligibility testing under laboratory conditions employing listener panels is often utilized. These laboratory evaluations often examine the system being tested under relatively "ideal" conditions. It would be more realistic to conduct laboratory intelligibility testing of voice communication systems with the systems and the listeners in as near an operational configuration as possible. For example, in evaluating an airborne radio system, the radio should be evaluated in conjunction with a standard aircraft intercommunication system, all listeners and speakers should wear standard custom fitted helmets, earphones and oxygen masks, through which they breathe compressed air and which contain a standard microphone. Likewise, the noise environment in which the systems are to be eventually used should be modelled as closely as possible. This is true not only for the listener, in order to evaluate possible masking effects on intelligibility, but also for the talker, since ambient noise conditions will often cause the talker to modify his/her vocal effort, thereby altering the acoustic content of the speech. A number of studies have concerned themselves with measuring the intelligibility of standardized speech materials in aircraft noise environments both to determine the speech interfering effects of aircraft flyovers (1,2) and to attempt to develop a valid methodology to assess aircrews' ability to understand aural communication in an operational environment (3,4,5). The studies reported here had a three-fold purpose, first to evaluate the effect of different levels of a simulated operational noise on the intelligibility of standardized speech materials as processed through representative USAF voice communication systems, second to gather base-line performance data on the ARC-164 radio system, and third to determine whether a systematic difference exists between the intelligibility of male and female talkers under the different noise conditions. The gathering of base-line data on the ARC-164 is important because this system will serve as the reference against which the performance of systems developed in the immediate future under programs for secure, jam-resistant voice communications will be evaluated. The question of whether there is any systematic difference in intelligibility under noise conditions between male and female talkers is of importance because the proportion of females in aircraft related job specialties will undoubtably continue to increase in the future, and because of the increasing probability that future aircraft will incorporate voice warning systems which may utilize female voices, either analog or synthesized.

METHOD

Approach

The comparative intelligibility of standardized test materials processed through representative Air Force communication systems was measured in the presence of varying levels of simulated operational noise. Volunteer listeners wearing standard inflight helmets and oxygen masks responded to the communication signals under the specified experimental conditions. Decrements in comparative intelligibility were attributed to the level of the simulated operational noise employed and the characteristics of the voice communication system being evaluated.

Subjects

Ten subjects, five male and five female, were employed in the present study. All were recruited from the general civilian population. They were paid at an hourly rate for their participation, with a cash bonus awarded when the subject completed all scheduled sessions. The hearing levels of all subjects were no greater than 15 dB at any standard audiometric test frequency from 500 to 6000 Hz. After each 4 hour experimental session, two of the subjects were selected to have their hearing tested to insure that the experimental noise conditions were not resulting in hearing threshold shift. In no case was a threshold shift found.

Facilities

This study employed the Voice Communication Research and Evaluation System (VOCRES) of the Aerospace Medical Research Laboratory (6), which was described in detail in a paper presented earlier in this conference. This system has the capability to realistically model the major acoustic factors experienced

by crew members that may adversely affect voice communications.

As was the case in the jamming studies reported the first day of this meeting, all subjects wore individually fitted HGU-26/P flight helmets with the H-154A earcup assembly. Each subject also wore a standard Air Force oxygen mask with the M-101 noise cancelling microphone. Compressed air was respired through A-19 Diluter Demand Pressure Breathing Regulators set at normal operation for all subjects during the talking and listening phases of the study.

The low power system of the acoustic environment simulation facility was used to simulate the cockpit noise environment of a tactical fighter aircraft. A pink noise source was shaped by a 1/3 octave band spectrum shaper so that the spectrum measured in the test space was representative of that produced by a typical jet fighter aircraft. The noise level in the chamber was measured for each of the four experimental conditions (ambient, 95 dB, 105 dB and 115 dB of simulated cockpit noise) at each octave band from 250 Hz to 4 kHz. The ambient noise condition has an overall level of about 79 dB, while the 95 dB, 105 dB and 115 dB conditions are representative respectively of an F-16 on final approach, F-16 at normal cruise, and an F-15 at 40,000 ft at Mach 1.96. The amount of attenuation provided by the helmets and earphones as was measured in earlier tests was calculated for each octave band (7) and subtracted from the chamber noise level. This provided an estimate of the noise level at the subjects' ears in octave bands. These estimates are shown in Table I.

OCTAVE BAND CENTER FREQUENCY	OVERALL NOISE LEVEL IN dB re 20µPa			
	79 (Ambient Level)	95	105	115
250	63.2	79.3	88.9	98.4
500	48.4	64.9	74.3	84.0
1000	41.4	62.7	72.1	82.2
2000	32.4	57.8	67.4	77.1
4000	24.0	49.4	59.3	69.4

TABLE I. Estimate of octave band noise level in dB at subjects' ears.

Measurement Instrument

This study employed a standardized measure of intelligibility, the Modified Rhyme Test (MRT) as developed by House, et al. (8) for assessing voice communication effectiveness. The materials used consist of six lists of 50 one-syllable words that are equivalent (lists) in intelligibility. Nine different randomizations of each of the six lists were used giving a total of 54 sets of test words. These test words were presented embedded within a carrier phrase that was the same for each item. The display at the talker's station provided the text which the talker read, the listeners' stations provided a choice of six rhyming key words from which the listeners selected the one he/she judged was correct by pressing the appropriate button. To compensate for correct answers obtained by guessing a correction factor (# incorrect/# possible choices - 1) was applied to the scores.

Experimental Procedure

Order of presentation of background noise levels was randomized and measurements were completed on one system before they were started on another. The order in which the systems were tested were (1) ARC-34, (2) ARC-164, (3) AIC-25, (4) ARC-150. Data was not gathered for the ARC-34 under ambient noise conditions in the test chamber, while it was for the other three systems.

Five of the ten subjects, three male and two female, were selected to serve as talkers. These individuals served as talkers and listeners in a "round-robin" fashion. The talker on any one list served as a listener on previous and subsequent lists. Subjects participated for 4 hours per day in experimental sessions of about 40 minutes followed by 15 minute rest periods. All ten subjects were run simultaneously within the VOCRES facility. The helmets and earphones worn by the subjects provided about 20 dB attenuation of the environmental noise.

RESULTS

Figure 1 summarizes the results from this study. Percent correct intelligibility (adjusted for guessing) for MRT words processed through the tested systems is shown as a function of the level of a
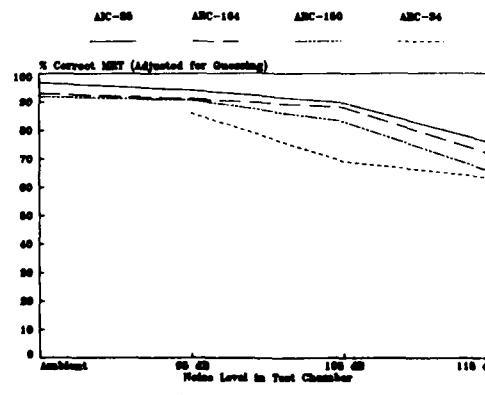

FIGURE 1

simulated operational noise environment. Data was also taken for all systems except the ARC-34 radio in the absence of a simulated operational noise, i.e., in the ambient noise of the test chamber. This ambient level was 79 dB. Examination of Figure 1 indicated that increasing the level of simulated cockpit noise in the test chamber resulted in decreased intelligibility for all four systems tested. The AIC-25 intercom alone showed a slight (7%) decrease in intelligibility as the noise level was increased up to 105 dB. A subsequent increase to 115 dB resulted in an additional 14% decrease in intelligibility or a score of 75% correct. When the speech test materials were processed through the selected radio systems in addition to the intercom, the effect of background noise level differed somewhat depending upon the system under test. The ARC-164, the current operational radio in US aircraft, yielded essentially the same function as did the intercom alone, except, that it was displaced downward two to five percent in intelligibility. The ARC-150, essentially an early version of the ARC-164 with fewer "slices" or interchangeable modules, performs the same as the ARC-164 in the ambient and 95 dB noise levels and five to seven percent poorer than the ARC-164 when the background noise is at 105 and 115 dB. The ARC-34, an older radio which was essentially replaced by the ARC-164, showed a quite large (17%) decrease in intelligibility when the background noise level increased from 95 to 105 dB. A subsequent increase to 115 dB resulted in an additional six percent loss in intelligibility, giving a score of 63 percent correct. It is also obvious that the shape of the function for the ARC-34 differs from that shown by the other three systems. This is possibly due in part to the difference in channel spacing between the radios. The ARC-34 has a channel spacing four times as wide as that of the ARC-164 and ARC-150 and therefore would accept much more random RF noise at the receiver. In addition, the frequency stability of the ARC-34 is much poorer than the other radios.

These results support the contention that laboratory intelligibility testing of voice communication systems should be conducted with the systems and the listening panel in as near an operational configuration as possible. The environmental noise in which talkers and listeners using the system will be operating should be considered an important variable. Not only does the noise environment affect the intelligibility of speech processed through a particular voice communication system, but the effects may differ between systems.

Male vs Female Speakers

Another question in which we were interested was whether the noise environment would have a differential effect on the intelligibility of male and female talkers. To address this question we looked at the data from one of the jamming studies reported earlier in this meeting. The data examined was that for the no jamming condition under different background noise levels (ambient, 95 dB, 105 dB, 115 dB). The noise spectrum was the same as reported above. The radio system used was the ARC-164. The data from the jamming study was used because that study contained the greatest number of observations at each data point. The results are summarized in Figure 2, where each data point is the average of 4 observations for each of 3 male speakers and 2 female talkers. Thus each data point is the average of 12 observations for the male talkers and 8 observations for the female talkers. The data for the AIC-25, ARC-34, ARC-150 and ARC-164 shown in Figure 1 was also examined in terms of male and female talkers and the data for each of these systems displayed the same trend as shown in Figure 2 for the jamming study data. It should be noted that four of the talkers (2 male and 2 female) and 5 of the listeners (3 male, 2 female) were different when the data in Figure 1 was collected from when the data in Figure 2 was collected.

Female vs Male Talkers

Male Talker   Female Talker


% Correct MRT (Adjusted for Guessing)
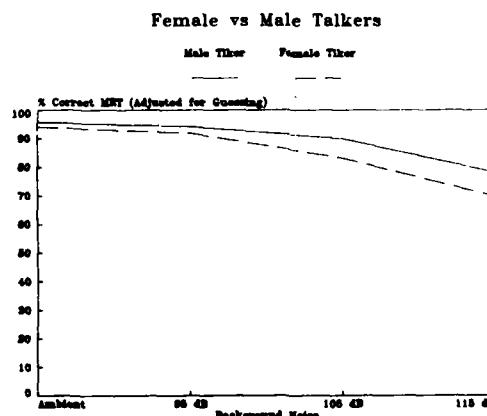
FIGURE 2

Examination of Figure 2 indicates that there was a systematic difference between the intelligibility of male and female talkers in the presence of simulated cockpit noise. In the ambient noise environment of the chamber and at 95 dB simulated cockpit noise there was little difference in intelligibility between male and female talkers (1.5 and 2.3 percent). At 105 dB background noise the male talkers are 6.8 percent more intelligible and at 115 dB this relative advantage has increased to 9.5 percent. These differences in relative intelligibility were tested for significance with a two-tailed t-test. The differences at 105 dB and 115 dB were found to be significant (p<.05).

Although these data indicate that as the level of simulated cockpit noise is increased female talkers become relatively less intelligible than male talkers, some care must be taken in generalizing these results. It is not known whether this effect is due to the generally higher-frequency content of the female voice being masked to a greater extent than a male voice as the background noise level is increased or whether the effect is due to the females modifying their vocal output as the noise level increased and thereby affecting their intelligibility. In order to separate out these effects studies will have to be

run where the talker and listeners are not in the same noise environment. Also it is not known whether this effect or its magnitude is specific to the type of cockpit noise simulated here or whether it would be different in another noise environment, e.g. transport aircraft, helicopter, ground station, etc. Because of the above noted expected increase in the number of females in aircraft related job specialties and the likelihood that voice warning systems will be increasingly incorporated into aircraft these questions should be addressed by future research.

SUMMARY

The results reported in this paper indicate that the presence of a simulated operational noise environment affected the intelligibility of standardized speech materials as processed through representative US Air Force voice communication systems and that the magnitude of the effect is not identical for all systems. In addition, evidence is provided that systematic differences exist between the intelligibility of male and female talkers under various levels of simulated cockpit noise, with female talkers being relatively less intelligible than male talkers as the noise level increased.

REFERENCES

1. Kryter, K.D. and Williams, C.E. "Masking of Speech by Aircraft Noise," J. Acoust. Soc. Amer., 39, 1966, 138-150.

2. Williams, C.E., Stevens, K.N., Hecker, M.H.L., Parsons, K.S. "The Speech Interference Effects of Aircraft Noise," FAA-TR-DS-67.19, Sep 1967, (AD660712).

3. Greene, J.W. "A Progress Report on the Naval Aviator's Speech Discrimination Test," Naval Aerospace Medical Research Laboratory Report NAMRL-1110, 1970, Pensacola, FL, USA.

4. Williams, C.E., Forstall, J.R. and Greene, J.W. "The Use of In-Flight Manikin Recordings for Evaluating the Communication Effectiveness of Flight Helmets," IEEE Trans. on Audio and Electroacous., Vol AU-19, #1, March 1971, 97-102.

5. Williams, C.E., Mosko, J.D., and Greene, J.W. "Evaluating the Ability of Aircrew Personnel to Hear Speech in their Operational Environments," Av. Space and Environ. Med., Feb 1976, 154-158.

6. McKinley, R.L. "Voice Communication Research and Evaluation System," AMRL-TR-80-25, 1980.

7. "Hazardous Noise Exposure," U.S. Air Force Regulation 161-35, July 1973.

8. House, A.S., Williams, C.E., Hecker, M.H.L., Kryter, K.D. "Articulation Testing Methods: Consonantal Differentiation with a Closed Response Set," J. Acoust. Soc. Amer., 37, 1965, 158-166.

DISCUSSION

MR. P.D. WHEELER (UK)

To what criterion (e.g., sidetone intelligibility or vu meter setting etc) do you instruct talkers to set their voice level during testing, as background noise level changes?

DR. T.J. MOORE (UNITED STATES)

The talker is instructed to monitor the vu meter mounted on his desk. The instructions are to maintain the vocal effort necessary to register a reading of zero on the vu meter when saying the words "Mark" and "Please". In the carrier phrase "You will mark the work.....please."

DR. K.E. MONEY (CANADA)

In the subjects who improved after 2 weeks off, was that a statistically significant improvement and if so, how could you account for that?

You do seem to get a novelty effect in this kind of research that improves performance.

DR. T.J. MOORE (UNITED STATES)

We did not run a statistic on that because we treated it as an exploratory pilot study and the limited number of subjects involved. If I had to account for it, I would say simply that, to the individuals it was novel again and that might be justified in the motivation being higher and after awhile it gets very old when constantly listening at this +3 dB V-S ratio. Now when we run the studies at the panel, the order of conditions is randomized, so on one trial they may get a no-jamming condition, the next might be +6 dB, the next one might be -3 dB, so it is not as tiring. That may reflect the fact that we selected one jammer, one level of jamming and just gave it to them trial after trial.

I would like to describe it as a motivational effect, which we make special efforts to control with the panel which in the exploratory study, we did not worry about that much.

SECOND LANGUAGE SPEECH COMPREHENSION IN NOISE   A HAZARD
TO AVIATION SAFETY

by

Hans M.Borchgrevink MD BA
Institute of Aviation Medicine
P.O.Box 281 Blindern, Oslo 3
Norway

SUMMARY

Simple Norwegian and English sentences were read by a bilingual adult, taperecorded
and presented individually to bilingual adults with English or Norwegian as their
first language and good command of the other. Each 65dB sentence was first presented
in so strong background USASI noise (75dB) that it could not be perceived, and was
repeated with the noise level progressively reduced in 2 dB steps from presentation
to presentation until the sentence was adequately repeated by the subject.

The results (table 1) demonstrated that for both groups the first (native) language
sentences were correctly repeated after fewer presentations, that is at a lower signal
- to - noise ratio, than the second language sentences. The difference between the
first (native) language comprehension threshold and the second language comprehension
threshold was statistically significant for both the Norwegian and the English subject
groups (p < 0.001 with t-test).

Individuals thus need fewer acoustical cues to comprehend sentences presented in their
first (native) language than corresponding sentences presented in their second language,
even when they have good command of their second language. Put in another way: The sub-
ject's ability to "fill in" or guess what is hidden in the background noise is better
in his native language. The phenomenon may influence aviation safety due to misinter-
pretation of messages received in noisy environment, for instance, if the pilot or air
traffic controller is experienced and the message is unexpected, and especially if one
or both of them are non-native speakers of English. This should be taken into account
in the selection and routine control of aircrew and airtraffic controllers - as well as
give highpriority to the development of improved intercommunication systems.

INTRODUCTION

Normal speech contains superfluous information, far more than essentially needed
for communicating the message. If the "informational overflow" is reduced below a
certain level, the communicated message will not be adequately comprehended by the
listener. This may for instance occur in case of poor signal-to-noise ratio (the speech
is masked by noise) or poor concept - reference coherence (the spoken words or sounds
means little or nothing to the listener and therefore do not evoke the appropriate
associations and concepts/ideas in the listener).

Modern aviation relies heavily upon verbal communication between aircrew and air
traffic controllers. The communication is mediated by intercommunication systems with
amazingly poor characteristics as regards signal-to-noise ratio and distortion.

English is the language used in aviation throughout the world, which means that a
considerable number of flights involve aircrew or air traffic controllers for whom
English is a foreign language. If one needs a better signal-to-noise ratio to compre-
hend messages presented in a foreign language, the aircrew and air traffic controllers
that are non-native speakers of English might accordingly represent a hazard to aviation
safety due to increased chance of misinterpretation or lack of comprehension essential
messages. The English language generally used in aviation is strongly influenced by
technical jargon, which should be familiar to all individuals trained for key positions
in aviation or air traffic control regardless of native language. This might counteract
and more or less eliminate an eventual negative effect of second language upon speech
comprehension in routine aviation.

However, for reasons of aviation safety, aircrew or aircrew traffic controllers
must be capable of understanding unexpected messages which most likely will occur in
critical situations. It was thus considered most relevant to approach the problem by
stating whether or not a person generally does need a better signal-to-noise ratio to
comprehend second language speech compared with first (native) language speech, merely
due to less established concept - reference coherence in the second language.
Using stimuli that were short, regular statements of the type noun + verb + complement/
adjunct, syntax and intonation as well as prosody would provide the listener with few
cues to sentence comprehension. If one presented taperecorded, isolated sentences to a
listener in his first and second language read by the same (truly) bilingual speaker,
the influence of cognitive capacity, experience, concept familiarity and nonverbal
communicational elements would be controlled, leaving concept - reference (word)
coherence in the first and second language as the variable.

In such a paradigm, the syntax of all sentences as well as the content/semantics
must be unknown to the subject, while all the words and concepts have to be known.
Further is needed a corresponding degree of concept and word familiarity in the two
different sets of sentences (one set in each of the two languages). This can be com-
pensatorily controlled by presenting the same two sets of sentences to two groups of

listeners, each being native speakers of each of the two languages and having <u>good</u> <u>command</u> of the other (second) language. If both groups of listeners then comprehended the sentences presented in their <u>first</u> language at a <u>lower</u> signal-to-noise ratio than the sentences presented in their second language, sentence intelligibility would be demonstrated to be a function of concept - reference coherence. In that case second language speech comprehension in noise would have to be regarded a true hazard to aviation safety due to increased chance of misinterpretation and consequent lack of comprehension of essential messages.

METHOD

The comprehension tresholds for sentences presented in the first (native) language and the second language were compared for 13 Norwegian adults (males and females) with university degrees in English, and for 13 English adults (males and females) with university degrees or corresponding command of Norwegian. The subjects were recruited from the University of Oslo, the British Council, Oslo and the British Embassy, Oslo (Norway). Normal hearing was confirmed by pure tone audiometry prior to testing.

The stimuli were 10 Norwegian sentences and 10 English sentences, all semantically different, but with corresponding syntax. They were short, everyday utterances of about four lexival words each, typically noun + verb + complement/adjunct. The sentences were devoid of specific cultural idioms. The phoneme frequencies (occurrence) in each sentence set corresponded to that of the respective language. The initial words in the sentences gave no clue to the completion. No thematic relationship existed between any of the sentences, which were all unknown to the subjects. The words included were familiar to everybody in both languages.

The sentences were read by a bilingual male adult as if occuring in daily conver- ation. His reading was supervised by university professors with native command of each language. The sentences were then tape-recorded on a full-track Nagra Kudelski tape- recorder (Bruel & Kjær condenser microphone), from which they were copied on to one track of a two-track Nagra Kudelski tape-recorder (on line), each sentence repeated 10 times with 6 secs. intermediate intervals. USASI-noise was recorded on the other track, starting about ½ sec. before each sentence repetition and ending about ½ sec. after the last word of the sentence. The sentence was thus "hidden" in the noise.

During testing the subject was sitting alone in an easy-chair in a soundattenuated room facing a loudspeaker which was mounted with the center 140 cm in front of his face (Tandberg TL 5010 loudspeaker, modified by the factory for optimal linearity). The investigator, placed in the adjoining room, could check the subject's position through a window. The two could communicate with each other through a separate intercommunication system. The subject was given a short standardised written instruction. When his understanding was verbally confirmed, both tracks of the tape-recorder were connected to the loudspeaker. Each sentence was first presented with so strong background noise (so low signal-to-noise ratio) that it could not be perceived, and was repeated with the USASI noise level progressively reduced in 2 dB steps from presentation to presentation. The sentences were always presented at 65 dB SPL (sound pressure level, dB rel. to $2.10^{-5}N/m^2$ measured free field in the test chamber just in front of the ear of a subject seated as described above). The noise level was 76 dB SPL at the first presentation of the sentence, being lowered (in 2 dB steps) to 56 dB SPL at the last presentation of each sentence. The task of the subject was to repeat the utterance verbally in the 6 sec. pause following each presentation. The investigator listened, marking how many presentations the subject needed to repeat each sentence correctly, also indicating the kind of error if the subject was close to the solution. The sentences in each of the two languages were presented in a fixed order and at the same sound pressure levels for all subjects. Half of the subjects in each native language group started with the English sentences, the other half started with the Norwegian sentences. They were informed whether the following sentence would be in English or in Norwegian.

RESULTS

The results demonstrated that the first (native) language sentences were correctly repeated after fewer presentations than the second language sentences, both by the English group of subjects and by the Norwegian group of subjects. The difference between the first language repetition threshold and the second language repetition threshold is statistically significant for both subject groups ($p < 0.001$) when applying t-test on the mean number of presentations needed for correct repetition of sentences (table 1a). The same is true when using a criterion that tolerates error in one lexical word <u>or</u> in two grammatical words (table 2a). Mean sentence repetition thresholds in terms of signal- to-noise ratio are given in tables 1b and 2b.

DISCUSSION

As all the subjects had normal hearing and the sentences were presented by tape- recorder, any recorded difference in sentence repetition faculty reflects the influence of non-acoustical parameters. Sentence intelligibility has been shown to decrease with decreasing contextual influence (Stowe & al. 1963). For instance, ungrammatical phrases with no meaning are more difficult to "shadow" than grammatical, meaningful sentences (Miller & Isard 1963). Repetition of nonsense sounds will accordingly require better signal-to-noise ratio than is needed for repetition of meaningful sentences. If one in- creases the signal-to-noise ratio of an unknown sentence, starting at an unintelligible level as in this experiment, the sentence comprehension threshold will accordingly be

reached <u>before</u> the "nonsense sound repetition threshold". Consequently, sentence repetition reflects sentence comprehension in the present paradigm.

The influence of cognitive capacity, experience, concept familiarity and communicational context were controlled by presenting the same two sets of tap-recorded Norwegian and English sentences to two groups of listeners, each having native command of one language and good non-native command of the other. The influence of grammatical knowledge was controlled by the use of short, regular phrases devoid of cultural idioms. The recorded difference between first and second language comprehension thresholds must therefore be ascribed to different degrees of concept - reference coherence/establishment in the listeners' first and second languages, following the reasoning presented in the introduction.

The results show that individuals need <u>fewer</u> acoustical cues to comprehend sentences presented in their first (native) language than corresponding sentences presented in their second language, even when they have good command of their second language. Put in another way - the subject's ability to "fill in", guess or synthesise what is hidden in the background noise, is better in his native language, presumably because his first language with better concept-reference coherence gives a better access to a righer association network in his single semantic memory (Borchgrevink 1981). This is in accordance with the "analysis by synthesis" model of speech perception (Stevens & House 1972). In this model the listener is believed to perform a progressive analysis of the utterance as it is pronounced by the speaker, trying to guess (synthesise) the most probable content of the entire utterance based on the analysed information available at each moment - consulting "lexicon stores", grammatical rules, and knowledge of the world acquired through previous experience.

Any percept must be regarded the brain's interpretation of a giver sensory input judged against a background of enviromment and experience. For the brain a percept can therefore not be false or correct - only the best (most probable?) solution when taking into account the actual circumstances of influence. A marginal signal-to-noise ratio will leave the listener with few cues to a presented verbal message, increasing the chance of the brain making a wrong "guess" and reducing the chance of the brain being able to realize that the guess was wrong. Consequently the brain may stick to the wrong guess, which will appear to be the correct message for the listener, who will behave according to the (wrong) message. In aviation such misinterpretation made by aircrew or air traffic controllers may easily lead to disaster.

The present experiment measures second language speech perception under optimal conditions: a truly bilingual person speaking to academically trained people with good (university level) command of their second language. Second language speech comprehension would be expected to be even poorer if the second language competence was lower in the listener. The same tendency would be expected if the speaker had a foreign accent with errors in phoneme pronunciation and prosodic features. On must therefore conclude that second language speech comprehension in noise must be regarded a hazard to aviation safety due to increased chance of misinterpretation and consequent lack of comprehension of verbal messages essential for flying.

Adequate prophylaxis would include 1) a reduction of noise and distortion in the intercommunication systems presently available, 2) the introduction of specific tests for speech comprehension in noise in the program for selection and routine examination of aircrew. Such tests are being developed at our institute (Borchgrevink & Andersen 1981). To assure adequate comprehension of the unexpected message that might occur in critical situations, we consider tests where verbal material other than routine flight orders are used as speech stimuli.

REFERENCES

Borchgrevink, H.M. (1981) "Speech perception involves cognition: the influence of concept - reference coherence upon speech perception",Submitted to Cortex.
Borchgrevink, H.M. & Andersen, H.T. (1981) "Sentence comprehension in noise is a test for aircrew and air traffic controllers", AGARD Specialists' Meeting on Aural Communication in Aviation, Soesterberg - . AGARD Conference Proceedings No.
Miller, G.A. & Isard, S. 1963. Some perceptual consequences of linguistic rules. <u>Journal of Verbal Learning and Verbal Behavior</u>, 2, 217-228.
Stevens, K.N. & House, A.S. 1972. Speech perception. In Tobias, J.V. (Ed.) <u>Foundations of Modern Auditory Theory</u>, vol. 2. London/ New York.
Stowe, A.N., Harris, W.P. & Hampton, D.B. 1963. Signal and context components of word-recognition behavior. <u>Journal of the Acoustical Society of America</u>, 35, 639-644.

TABLE 1   CORRECT REPETITION OF SENTENCES

a) number of presentations	Norwegian subjects(n=13)		English subjects(n=13)		T-test	
	mean	SD	mean	SD	value	prob.
10 Norwegian sentences	6.55	0.64	8.55	1.33	4.88	p < 0.001
10 English sentences	7.54	0.83	6.31	0.68	4.13	p < 0.001
b) signal-to-noise ratio needed (dB)						
10 Norwegian sentences	0.1	1.28	3.9	2.66	4.88	p < 0.001
10 English sentences	2.1	1.66	0.4	1.36	4.13	p < 0.001

TABLE 2   SENTENCE REPETITION WITH ERROR IN ONE LEXICAL WORD
OR IN TWO GRAMMATICAL WORDS

a) number of presentations needed	Norwegian subjects(n=13)		English subjects(n=13)		T-test	
	mean	SD	mean	SD	value	prob.
10 Norwegian sentences	5.22	0.61	6.98	1.26	4.53	p < 0.001
10 English sentences	6.60	0.77	5.02	0.63	5.73	p < 0.001
b) signal-to-noise ratio needed (dB)						
10 Norwegian sentences	-2.6	1.22	1.0	2.52	4.53	p < 0.001
10 English sentences	0.2	1.54	-3.0	1.26	5.73	p < 0.001

DISCUSSION

DR. J.D. MOSKO (UNITED STATES)

Have you had the opportunity to determine which feature (i.e., nasality, voicing, etc) emerged sequentially from the noise for each language?

DR. H.M. BORCHGREVINK (NORWAY)

No. To avoid that certain keywords in the sentences were heavily stressed, we looked at the sound pressure level curves for each sentence and identified which parts of the sentence that was responsible for which peak, but our analysis was not brought further than that.

As mentioned under METHOD, the listeners were told beforehand the language of the following sentences, as I consider it to be the case in most situations that you know in which language a speaker will address you.

MR. R. PLOMP (NETHERLANDS)

In your procedure, sentences seem to be always presented first at a below-threshold signal-to-noise ratio and then this ratio is improved in steps of 2 dB and the sentence repeated until it is understood correctly. In our experience repetition of the same sentence may introduce extra spread in the data because if a sentence is understood incorrectly by the listener, it is rather difficult for him to get rid of this bias even for a favourable signal-to-noise ratio.

DR. H.M. BORCHGREVINK (NORWAY)

Some very few subjects behaved according to your description and could be "stuck with" an inappropriate comprehension of the sentence as you point out. However, this did not happen often, which is illustrated by the small standard deviation demonstrated for comprehension of first-language sentences. The somewhat larger standard deviation for second language speech comprehension threshold (in noise) in my opinion, reflects inadequate "fill in" due to poor concept-reference establishment in the second language rather than not getting rid of an inappropriate interpretation of the sentence.

DR. K.E. MONEY (CANADA)

When you use an inflight test of hearing,

(i)   Do you allow the pilot to select the type of aircraft, to get an aircraft with a good signal-to-noise ratio?
(ii) Do you allow the pilot to set the sound intensity to his headset wherever he wants it, or do you set it in a standard way?

DR. H.M. BORCHGREVINK (NORWAY)

So far the testing of the pilot's hearing during actual flight has been performed locally and with no restrictions as to e.g., selection of aircraft. The purpose of such testing has been to make sure whether a pilot or air traffic controller with a hearing loss approaching the qualification criterion, is still capable of adequate comprehension of verbal messages received through radio to headset in noisy conditions. The selection of aircraft and other devices has been left to the local commander. Of course, this gives the pilot the opportunity to be tested under optimal conditions.

No restrictions. However, we do inform the pilots and air traffic controllers about the risk of getting a hearing loss from high sound intensity in his headset. A high speech intensity level may be beneficial in some cases, but hardly for persons with high frequency hearing loss because too big amplifications in the lower frequencies appears to increase the critical bandwidth.

THE EFFECTS OF EAR PROTECTORS AND HEARING LOSSES ON SENTENCE INTELLIGIBILITY IN AIRCRAFT NOISE

by

G.R. Fröhlich, Col., GAF, MC
German Air Force Institute of Aviation Medicine
P.O. Box 172 KFL, D-8080 Fürstenfeldbruck, Germany

INTRODUCTION

In 1946 KRYTER has already stated, that earplugs improve speech intelligibility in higher noise levels. This statement has been endorsed by later investigators mainly on normal hearing subjects. Contrary to this, flight line personnel with hearing defects often complain, that face-to-face speech communication in noise is considerably reduced when ear protectors are worn. The aim of our study was to determine whether this could be confirmed or not. An effective noise protecting flight helmet changes the flat jet aircraft cabin noise spectrum into a spectrum with predominance of lower frequencies. Thus, our paper may also contribute to the question, whether the additional wearing of earplugs under the ear cups might improve speech perception.

METHODS

As test material we have used the Marburger Sentence Intelligibility Test (NIEMAYER), which consists of 8 phonetically balanced sets of 10 German 5-word-sentences. Thus each set consists of 50 words and every correctly repeated word has a score of 2%, for example:
1. Geld allein macht nicht glücklich.
2. Böse Menschen verdienen ihre Strafe.
As background noise we have used a predominantly low frequency aircraft cockpit noise (Table 1) with an AOSPL of 94 dB lin and 86 dB A, the Speech Interference Level (SIL) centered at 500, 1000 and 2000 Hz was 78 dB.

Table 1: Spectrum of Aircraft Cabin Noise

0,25	0,5	1,0	2,0	4,0	8,0	kHz
93	87	77	71	64	53	dB

We have investigated the speech intelligibility in noise of 3 different groups according to Table 2 and Fig. 1-3. Every subject was tested under 3 different conditions:
a) Without ear protectors
b) Wearing the low pass filter type earplug Selectone K
c) Wearing the Willson SB 258 ear muff with excellent sound attenuation already at medium frequencies.
The 78 SIL background noise was presented together with sentence speech levels of 75, 80, 85,90 and 95 dB until 100 % intelligibility was scored.

Table 2: Percent Intelligibility Scores Obtained at Various Speech Levels in Noise (Means and Standard Deviations)

Category of Hearing	Age		75 dB	80 dB	85 dB	90 dB	95 dB	100 dB
Normal Hearing (N=20)	22	no Protection	$47^{\pm}21$	$84^{\pm}12$	100			
		with Plug	$46^{\pm}30$	$90^{\pm}11$	100			
		with Muff	$59^{\pm}32$	$95^{\pm} 9$	100			
Hearing Losses above 2000 Hz (N=20)	41	no Protection	$20^{\pm}17$	$65^{\pm}19$	$88^{\pm} 9$	$98^{\pm} 4$	100	
		with Plug		$37^{\pm}28$	$75^{\pm}23$	$94^{\pm} 7$	100	
		with Muff		$42^{\pm}21$	$83^{\pm}18$	$97^{\pm} 5$	100	
Substandard Hearing at 1000 and 2000 Hz (N=14)	45	no Protection	$43^{\pm}23$	$69^{\pm}20$	$89^{\pm}10$	$97^{\pm} 7$	100	
		with Plug	$14^{\pm} 8$	$49^{\pm}23$	$74^{\pm}25$	$88^{\pm}19$	$94^{\pm}13$	
		with Muff		$29^{\pm}18$	$51^{\pm}20$	$76^{\pm}23$	$90^{\pm}11$	$97^{\pm} 3$

In Fig. 1-3 the right side shows the pure tone audiograms and the additional hearing losses due to ear protectors. The kidney-shaped speech area is calculated for normal conversational voice 2 cm from the speaker's lips. The lightly shaded part of it is masked by the aircraft noise, only the heavily shaded area is available for speech perception. The left side shows the sentence discrimination at various speech levels related to the pure tone audiograms with and without ear protectors.

RESULTS

Normal hearing subjects (Fig. 1) without ear protectors score 84% at 80 dB and 100% at 85 dB speech level. With the ear plug and with the ear muff there is a slight increase in scores, thus confirming previous findings.

Subjects with high tone hearing losses (Fig. 2) score 65% at 80 dB, 88% at 85 dB and 98% at 90 dB speech level. With plugs and muffs there is a marked decrease in scores at 80 dB and to a lesser degree at higher speech levels.

Subjects with substandard hearing in the speech area 500-2000 Hz (Fig. 3) have the same scores as the previous group. With the low pass filter ear plug there is a 20% decrease of intelligibility at all speech levels. 100% could not be obtained even at 95 dB. This negative effect is even more pronounced with the ear muff because there is a decrease of 40% at low and moderate signal-to-noise ratios. A 100% score could not be obtained even at 100 dB speech level.
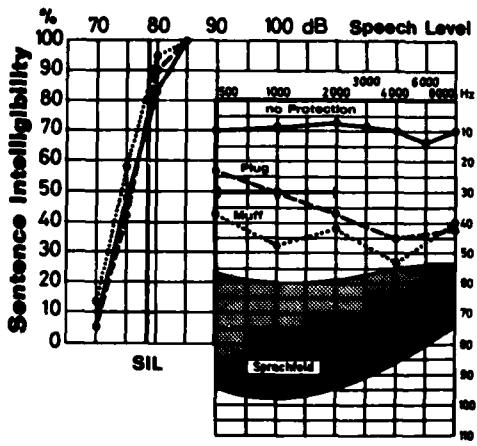

Fig. 1: Normal Hearing Subjects
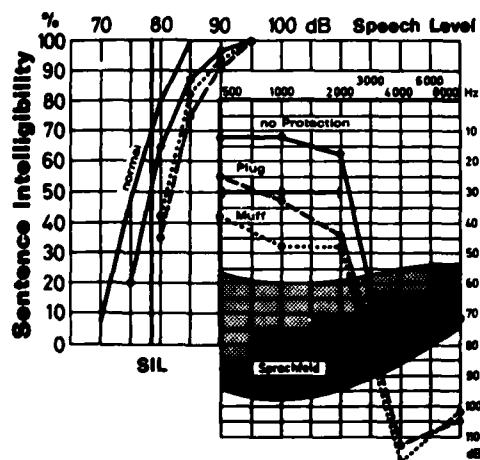
Fig. 2: Hearing Losses Above 2000 Hz
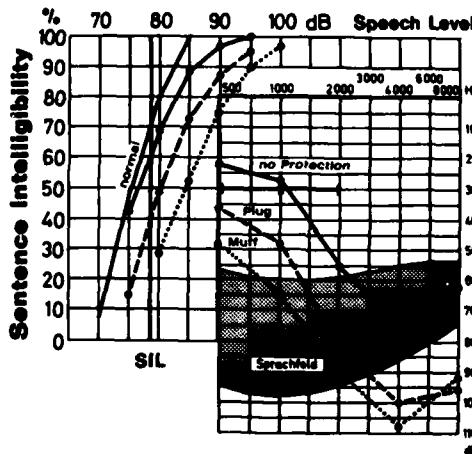
Fig. 3: Substandard Hearing

DISCUSSION

For normal hearing subjects the lower speech frequencies up to 500 Hz are masked by the noise, but the remaining medium and higher frequencies are sufficient to warrant adequate speech intelligibility, though the levels are at the threshold of discomfort. The use of ear protectors creates an artificial conductive-deafness which reduces high noise and speech levels to a comfortable level and thus slightly improves speech intelligibility.

For subjects with marked high tone hearing losses, the situation is different. Not only the lower speech frequencies are masked by the noise, but also the higher speech frequencies cannot be perceived due to the hearing defect. The remaining medium frequencies are insufficient to warrant good speech intelligibility at low signal-to-noise ratios (Cocktail Party-Effect). Ear protectors cut off additional middle and higher speech frequencies and thus further decrease speech perception.

Sensoryneural hearing losses in the main speech area as presented in Fig. 3, do not completely cut off higher speech frequencies, therefore with the help of recruitment intelligibility is rather satisfactory. However, if we add the sound attenuation of an effective ear protector, we produce a mixed deafness so severe, that recruitment no longer can be effective and natural voice, even at its highest vocal effort, cannot be properly perceived.

CONCLUSION

The complaints of deaf flight line personnel that ear protectors reduce speech communication in higher noise levels could be confirmed. This may also pertain to wearing ear plugs beneath ear cups.

STI - AN OBJECTIVE MEASURE FOR THE PERFORMANCE OF VOICE COMMUNICATION SYSTEMS

T. Houtgast and H.J.M. Steeneken

Institute of Perception TNO, P.O. Box 23
3769 ZG Soesterberg, The Netherlands

SUMMARY

A measuring device has been developed for determining the quality of speech communication systems. It comprises two parts, a signal source which replaces the talker, producing an artificial speech-like signal, and an analysis part which replaces the listener, by which the signal at the receiving end of the system under test is evaluated. Each single measurement results in an index (ranging from 0-100%) which indicates the effect of that communication system on speech intelligibility. The index is called STI (Speech Transmission Index). A careful design of the characteristics of the test signal and of the type of signal analysis makes the present approach widely applicable. It has been verified experimentally that a given STI implies a given effect on speech intelligibility, irrespective of the nature of the actual disturbance (noise interference, band-pass limiting, peak clipping etc.).

1. INTRODUCTION

A common method for determining the quality of speech transmission channels is by performing intelligibility tests with talkers and listeners using sentences, rhyme words, or other test material. This approach has the obvious advantage of its directness. However, there are some serious drawbacks, such as the need of a great number of trained talkers and listeners and the poor information on the type of degradation on the channel.

This has led to the development of objective measuring techniques. French and Steinberg (1) published a method for predicting the speech intelligibility of a transmission channel from its physical parameters. By using this method a relevant index (Articulation Index, AI) was obtained. The method was reconsidered by Kryter (2) who greatly increased its accessibility by the introduction of a calculation scheme, work sheets, and tables.

The AI method is particularly appropriate for channels with distortions such as interfering noise and band-pass limiting, and is not simply applicable when nonlinear distortions (digital signal processing, peak clipping) or distortions in the time domain (reverberation, echoes) are involved.

The present approach is more generally applicable as it also accounts for disturbances in the time domain and for nonlinear distortions.

The channel under test is considered as a **black box**, from which the speech degradation per octave band is determined by means of measurements performed with a special test signal.

Due to the specific nature of the test signal, reflecting the spectral and temporal characteristics of running speech, the method accounts correctly for a wide variety of distortions.

The theory underlying the characteristics of the test signal and the index to be derived (the STI, Speech Transmission Index) will be mentioned briefly.

2. SPEECH ENVELOPE SPECTRUM

Generally speaking, the information in a signal is carried by its temporal level fluctuations. In view of the ear's frequency selectivity, this should be formulated more precisely: the information is carried by the fluctuations of a signal's envelope within each of a series of parallel frequency bands.

An objective index related to speech intelligibility may be based on the extent to which the original fluctuations are preserved in the signal reaching the listener. This can be based on an analysis of the envelopes within octave bands. This degree of frequency selectivity may seem crude in relation to hearing, but recognises the fact that in most practical conditions a possible frequency-dependency of the interference need not be considered more precisely than in octave intervals.

An objective quantification of the amount of fluctuations present in a given signal is illustrated in Fig. 1. In the upper part of this figure the intensity as a function of time is plotted. The dotted line represents the average intensity. Many things may affect the envelope function. For instance, constant interfering noise will raise the average intensity and, as a consequence, the **relative** fluctuations (in terms of modulation depth) will decrease. A quantitative description of the fluctuations of the envelope function is provided by its frequency spectrum, the so-called **envelope spectrum**. An illustration of such an envelope spectrum, ranging from 0.25 Hz up to 25 Hz, is given in the lower part of Fig. 1. The envelope spectrum indicates to what extent the different fluctuation

RUNNING SPEECH
ENVELOPE
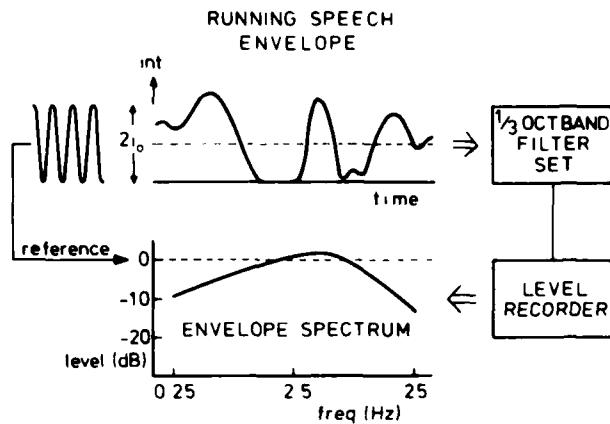
Fig. 1 Illustration of the enve-
lope function and envelope spec-
trum of running speech.

rates are present in the envelope function. I is normalized with respect to the avarage
intensity of the envelope function.

Now, after this introduction of the envelope spectrum, what does it actually mean?
It is important to note that, as a result of the normalizing procedure, the envelope spec-
trum is independent from the level of the running speech analyzed. Let us consider some
simple characteristics of the envelope spectrum. A value of 0 dB for a particular 1/3-oc-
tave band indicates that the intensity is 100% modulated by modulation frequencies within
the corresponding 1/3-octave band, a value of -10 dB indicates 32% modulation, a value of
-20 dB 10% modulation, etc. When a signal with a constant level, as for instance a sine-
wave, is applied instead of running speech, the envelope spectrum is far down; when white
noise is applied the envelope spectrum is determined by the statistical envelope fluctua-
tions of octave-band filtered white noise.

The envelope spectrum thus defined quantifies the degree of modulation of the
speech envelope as a function of modulation frequency. Stable and reproducible results
have been obtained by basing the analysis on 40 sec running-speech samples (Houtgast and
Steeneken, 3). An example is presented in Fig. 2, representing the spectral distribution
of the envelope modulations encountered in running speech. In view of the notion that the
information in a signal is carried by its envelope modulations, it should prove interest-
ing to relate speech intelligibility to the degradation of the speech-envelope spectrum.

Fig. 2 The fluctuations of the envelope of
octave-band filtered speech (top panel) are
quantified in the form of a 1/3-octave band
spectrum normalized with respect to the mean
value $\bar{I}$. In this way the ordinate of this
envelope spectrum can be interpreted as rel-
ative modulation depth.

sample of running speech intensity
(within the 2kHz oct. band)



one-sec. intervals

1/3-octave band centre frequency
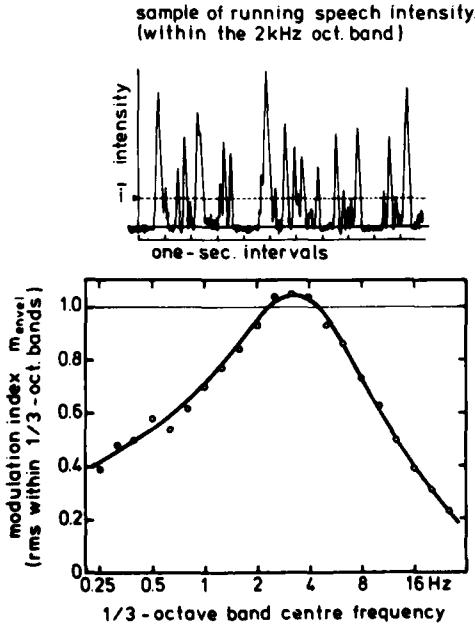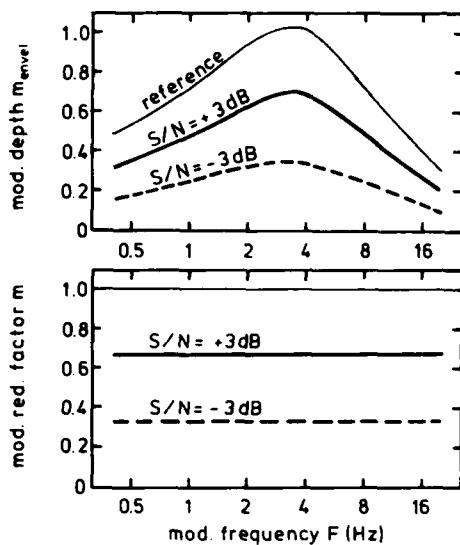
Fig. 3 Top panel: Envelope spectra for two conditions of steady-state interfering noise (the reference curve is taken from Fig. 2). Lower panel: The corresponding modulation transfer functions.

3. THE MODULATION TRANSFER FUNCTION (MTF)

As an illustration, the influence of a simple type of interference on the envelope spectrum will be considered: Steady-state interfering noise.

The effect of interfering noise on the original envelope function is simply an addition of a constant, equal to the mean noise intensity (within the octave-band considered). Since the envelope spectrum reflects the relative modulation depth (normalized now with respect to the new mean $\bar{i} + i_{noise}$), this implies a reduction of the original envelope spectrum by a constant factor, depending on S/N ratio only. In general, the factor m by which the original modulation depth reduces, as a function of modulation frequency F, will be named the **modulation transfer function**. Hence, in this case the modulation transfer function m(F) for noise interference is independent of F, and is given in the bottom panel of Fig. 3.

It is important to note that the modulation transfer function m(F) is a general characteristic of a given condition, which applies to any signal, either speech or any other complex or simple envelope function. This very feature allows a simple direct measurement of the modulation transfer function in actual conditions.
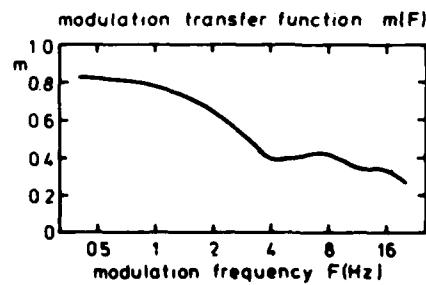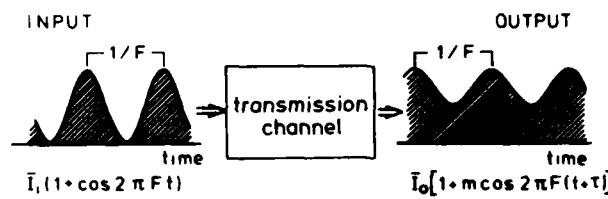

Fig. 4 Illustration of the principle underlying a direct measurement of the modulation transfer function.

The function m(F) quantifies the reduction of the envelope spectrum as caused by the speech communication system. The main feature of the measuring method is that the function m(F) is determined for each of a number of modulation frequencies individually. For that purpose the speech signal is replaced by a test signal containing a noise carrier which is 100% intensity modulated with variable modulation frequency F. An important requirement (in relation to the effect of interfering noise) is the adjustment of the long-term mean intensity of the test signal to that of the speech signal normally applied, for each individual octave-band considered. The resulting modulation depth m of the received signal directly represents the modulation reduction factor m(F) for that particular modulation frequency, as indicated in Fig. 4. By varying the octave-band applied to the received signal, the function m(F) can be determined for a number of individual octave-bands.

The modulation transfer functions thus obtained quantify the degradation of the original modulation of the (octave-band specific) speech envelope. Before investigating the relation between this function and the degradation of speech intelligibility, we will first consider some complications arising in case of nonlinear distortions.

4. CONSIDERING NONLINEAR DISTORTIONS

For transmission channels with only linear types of distortion (e.g., interfering noise and band-pass limiting), the test signal can be modulated simultaneously within all octave-bands, because the octave-bands do not interact. However, channels with a nonlinear transfer (e.g., peak clipping and digital modulation techniques) introduce harmonics and intermodulation components in other octave-bands, and therefore require a more sophisticated test signal. The more representative test signal would be running speech, in which only the speech signal ... octave-band being tested is replaced by the sine-modulated test signal. In this way representative disturbing components are introduced in the octave-band being tested, **uncorrelated** with the sinusoidal intensity modulation of the test signal within that octave-band.

Instead of running speech, the octave-bands not being tested may contain artificial speech-like signals. This is accomplished by switching on and off pseudorandomly the signal within each octave-band not under test. Physically, the envelope of each switched octave-band has been made similar to the envelope of running speech. This is performed by matching the fluctuation patterns (envelope spectra).
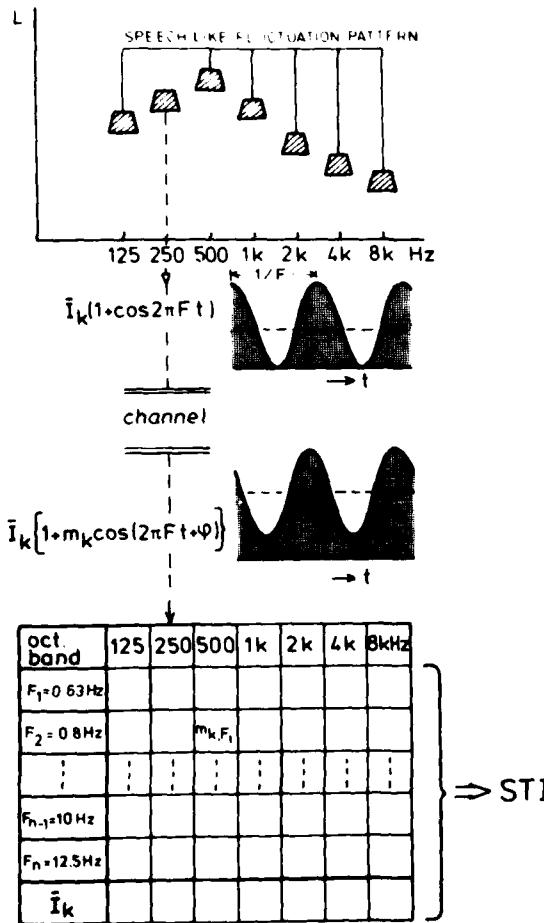

Fig. 5 Schematic description of the measuring procedure for obtaining the STI. This figure applies to the actual measurement for the octave-band with center frequency 250 Hz.

This leads to a measuring procedure as illustrated in Fig. 5. Seven noise bands with a width of 2/3-octave each (to avoid cross talk on the analysis side) are either intensity modulated with a cosine function or switched on/off pseudorandomly. For each octave-band the degradation of the test signal is measured by means of the reduction of the modulation index m for 14 modulation frequencies. For nonlinear transmission channels representative distortion components are introduced by the pseudorandom fluctuating test signal in the octave-bands not under test. The rms level of the cosine-modulated test signal within an octave-band must be equal to the corresponding long-term rms octave-band level of the speech normally applied to the channel under test. The rms level of the pseudorandomly modulated signal within each octave-band must be 3 dB higher than this target level. The data matrix thus obtained should be converted into one single index (the Speech Transmission Index, STI), indicating the effect of the channel under test on speech intelligibility.

5. STI AND SPEECH INTELLIGIBILITY

The procedure for deriving the STI from the data matrix as indicated in Fig. 5, is described in detail elsewhere (Steeneken and Houtgast, 4). Basically, each measured m-value is transformed into an **equivalent** signal-to-noise ratio, by using the theoretical relation between m-value and S/N ratio in case of actual steady-state noise interference (see Fig. 3). The weighted mean of these S/N ratios forms the basis for the STI.

We will now consider the relevance of the STI with respect to speech intelligibility. For that purpose a wide variety of 167 conditions were simulated and subjected to intelligibility measurements with talkers and listeners as well as to modulation transfer function analysis. The conditions refer to speech communication systems subjected to various degrees of interfering noise, band-pass limiting, peak clipping, automatic gain control, reverberation and combinations of these distortions.

The intelligibility scores were obtained with phonetically-balanced CVC-type nonsense words (PB-words), with a panel of 4 talkers and 5 listeners. Each condition is qualified subjectively by its mean PB-word score. It is well recognized that PB-word scores have only a limited value as an **absolute** quantification of intelligibility; the figure obtained may depend on the specific word material applied and on the skill of the talkers and the listeners. However, within this set of 167 conditions, each figure is meaningful as a relative qualification of intelligibility. The relation between STI and PB-word score for all conditions investigated is given in Fig. 6 together with the best-fitting curve based on all data points. The standard deviation is σ = 5.6%. It should be noted that the PB-word score measurements were performed with test words phonetically balanced for the **Dutch** language, but we feel that the conclusions obtained are applicable to other languages.

It should be realized that the STI method does not account for any frequency shift and frequency multiplication in the transmission channel under test. Several types of vocoder systems may also fall outside the scope of this approach. Some examples are given elsewhere (Steeneken and Houtgast, 5).
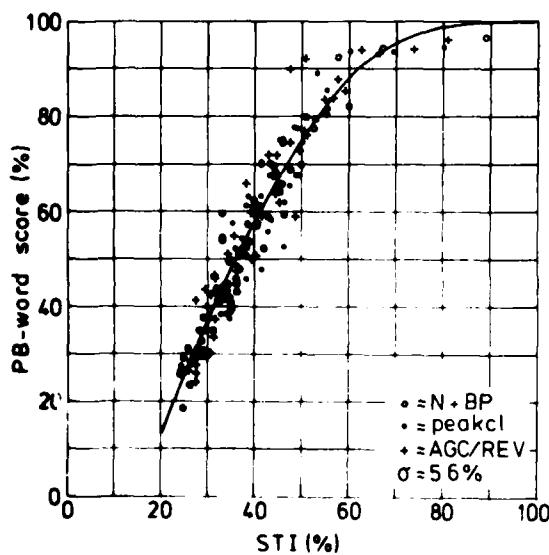

Fig. 6 Relation between STI and PB-word score for the conditions with noise, band-pass limiting, peak clipping, automatic gain control, and reverberation. The curve represents the best-fitting curve for all these data points.

6. CONCLUSIONS

A measuring procedure has been developed which is based on the observed reduction of the modulation index at the output of a speech transmission channel. Using a sophisticated test signal and calculation scheme, the resulting index STI was found to correlate well with intelligibility scores for a wide variety of speech communication channels.

A total number of 167 channels was considered, including distortions like band-pass limiting, noise, peak clipping, automatic gain control, and reverberation. For these channels both STI measurements and PB-word intelligibility scores were determined. The results showed a standard deviation of $\sigma = 5.6\%$ between actual intelligibility scores and those predicted by the STI values.

REFERENCES

(1) French, N.R. and Steinberg, J.C., Factors Governing the Intelligibility of Speech Sounds, J. Acoust. Soc. Am. 19, 1947, 90-119.

(2) Kryter, K.D., Methods for the Calculation and Use of the Articulation Index, J. Acoust. Soc. Am. 34, 1962, 1689-1697.

(3) Houtgast, T. and Steeneken, H.J.M., Envelope Spectrum and Intelligibility of Speech in Enclosures. IEEE-AFCRL 1972 Speech Conference.

(4) Steeneken, H.J.M. and Houtgast, T., A physical Method for Meauring Speech Transmission Quality, J. Acoust. Soc. Am. 67, 1980, 318-326.

(5) Steeneken, H.J.M. and Houtgast, T., Some Applications of the STI-Method in Evaluating Speech Transmission Channels. These proceedings.

SOME APPLICATIONS OF THE STI-METHOD IN EVALUATING SPEECH TRANSMISSION CHANNELS

Herman J.M. Steeneken and Tammo Houtgast

Institute for Perception TNO, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands

SUMMARY

A description is given of a measuring device for the application of the STI-method as described in the previous contribution by Houtgast and Steeneken. The application of the device in evaluating speech communication channels as radio communication links, digital communication channels and microphones and telephones in noisy environments is demonstrated. Consequently the STI might well be used as a design specification for speech communication systems.

1. INTRODUCTION

The STI-method as described in the previous contribution (1) "STI" – an Objective Measure for the Performance of Voice Communication Systems", is realized in a measuring device. In principle the method measures the Modulation Transfer Function (MTF) for seven relevant octave bands (125 Hz – 8 KHz). From the MTF the signal-to-noise ratio for each modulation frequency and each octave band can be calculated. From these signal-to-noise ratios a single Index (Speech Transmission Index, STI) is obtained which is a good predictor of the transmission quality (Steeneken and Houtgast (2)). The practical realization of a measuring device for measuring the MTFs consists of three parts: a test-signal generator, an analysis part and a micro processor system for control and calculations. In this paper we will restrict ourselves to a brief description of the generator and analysis part, a more extensive description is given by Steeneken and Agterhuis (3).

The generator part as given in Fig. 1 consists of a noise source followed by seven modulators. The outputs of the amplitude modulators are connected to a set of 2/3 octave filters. After this filtering the test signal is composed by adding the output signals of the filters in a summing amplifier. The contribution of each of the seven filter outputs is such that a spectral weighting had to be applied in order to obtain a speech-like spectrum. The modulation signal of the seven independent amplitude modulators for the seven channels can be switched either to random fluctuations of the type of the envelope function of running speech or to modulation by a Digital-to-Analog Converter (DAC). This DAC provides (under computer control) the sinusoidal intensity modulation of the channel being tested. The selection of the modulation mode is performed by a digital interface under program control. A synchronization-signal generator is included in the generation module. It generates on request a 1.0 or 0.5 sec FSK modulated signal (carrier 1166 Hz modulation 64 Hz). At the analysis side this signal can be detected and is used to start the measuring procedure especially for those conditions where generator and analyzer are separated (or when the test signal is prerecorded on tape). This synchronization signal has turned out to be detectable even on very poor communication channels.

The signal flow in the analysis module, as given in Fig. 2, starts with a line amplifier followed by a set of octave band filters and a computer controlled switch to select the frequency band (filter) under test. Amplification of the filtered signal is performed in an autorange amplifier. The gain of this amplifier is set by the computer I/O bus. After this amplification the envelope function of the received signal is derived by an envelope detector and low pass filter. This envelope is sampled by an analog-to-digital converter and processed by the computer.
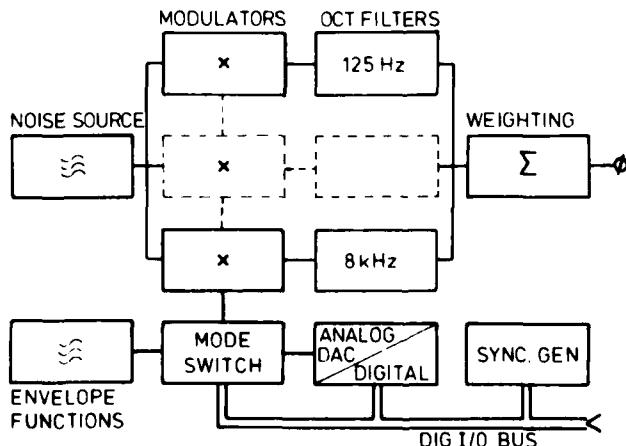

Fig. 1 Block diagram of the test signal generator.
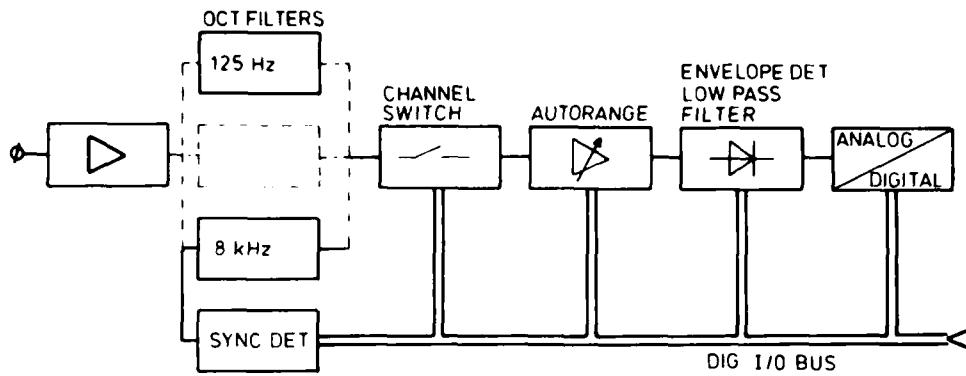
Fig. 2 Block diagram of the analysis part.

The synchronization signal detector is equipped with a phase-locked loop circuit to detect the 64 Hz FSK modulation. When the duration of a detected burst meets the requirements of 1.0 - 0.5 sec the device gives a digital output code to start a measurement (1 sec sync) or to proceed with the next octave band (0.5 sec sync). After completion of the measurements for the seven octaves the STI is calculated and displayed.

The aim of this paper is to demonstrate the application of the method on communication channels, therefore we will proceed in the next two chapters by giving some results of the validation of the system on wave-form coders and the evaluation of a Deltamodulation communication system and a telephone/microphone communication system.

2. VALIDATION OF THE STI-METHOD ON WAVE-FORM CODERS

For the validation of the STI-method on wave-form coders four different systems were used in the study:

- µlaw PCM system wordlength 8 bit, bitrate 64 Kbit/sec (PCM 1)
- PCM system wordlength 6 bit, bitrate 48 Kbit/sec (PCM 2)
- Deltamodulator, bitrate 9.6 Kbit/sec (CVSD 1)
- Continuous Variable Slope Deltamod., bitrate 32 Kbit/sec (CVSD 2)

With these basic systems 42 different transmission channels with a different performance were composed by adding bit errors in the link between coder/decoder and by adding noise to the speech at the analog input of the coders. For these various conditions the

Fig. 3 Relation between PB-word score and STI for 42 transmission channels with different wave-form coders. The solid line represents the best-fitting curve between the data points.
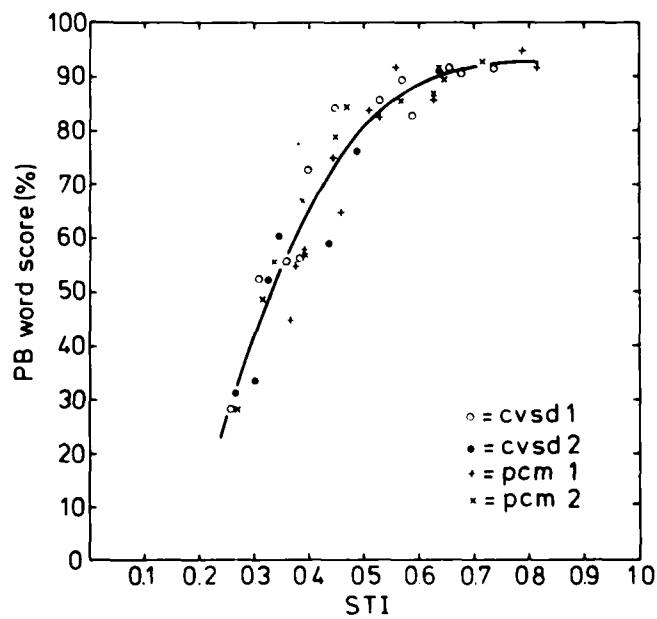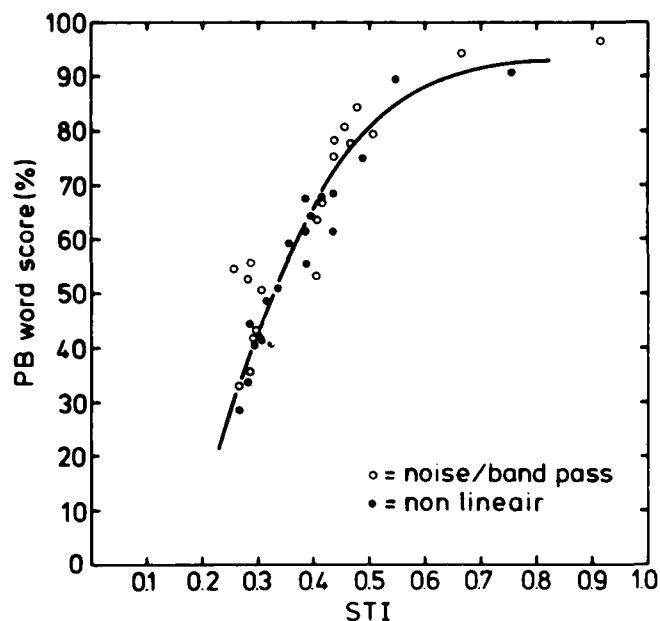
Fig. 4 Relation between PB-word scores and STI for 39 analog transmission channels. The solid line gives the best-fitting between the data points of Fig. 3.

subjective intelligibility as well as the STI-value was measured. The subjective measurements were carried out with monosyllabic nonsense words of the type consonant-vowel-consonant. Lists of 50 words, phonetically balanced (PB) for the Dutch language were used, for four talkers. At the receiving side 5 listeners were used. This means that each condition was tested by a 1000 nonsense words. Fig. 3 gives the relation between the STI-value and the PB-word score for the 42 conditions. The predictive power of the STI for the subjective score is expressed by the vertical spread of the data points around the best-fitting curve. The standard deviation of this spread is 5.5% which is of the same order as found before for analog systems. A point of interest was whether the relation between STI and PB-word score for the investigated digital system corresponds to the relation found before for analog systems. In other words, can we apply the same criteria expressed in STI for analog channels as well as for wave-form coders.

Therefore the same type of subjective and objective measurements was performed on a subset of 39 conditions as was studied before on analog channels. Fig. 4 gives the relation between STI and PB-word score around the best-fitting curve of Fig. 3. The figure shows that there are no systematic differences which is shown by the spread of the data points around this curve (expressed in σ = 7.7%).

Fig. 5 STI as a function of the bit error rate between coder and decoder for two types of Deltamodulators.
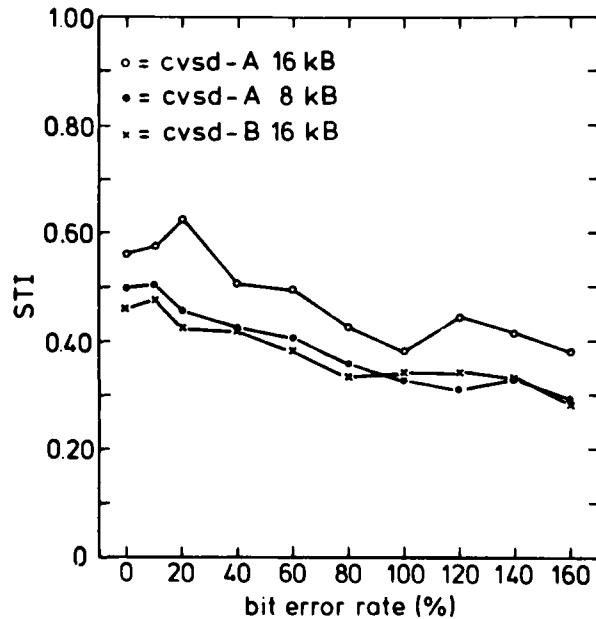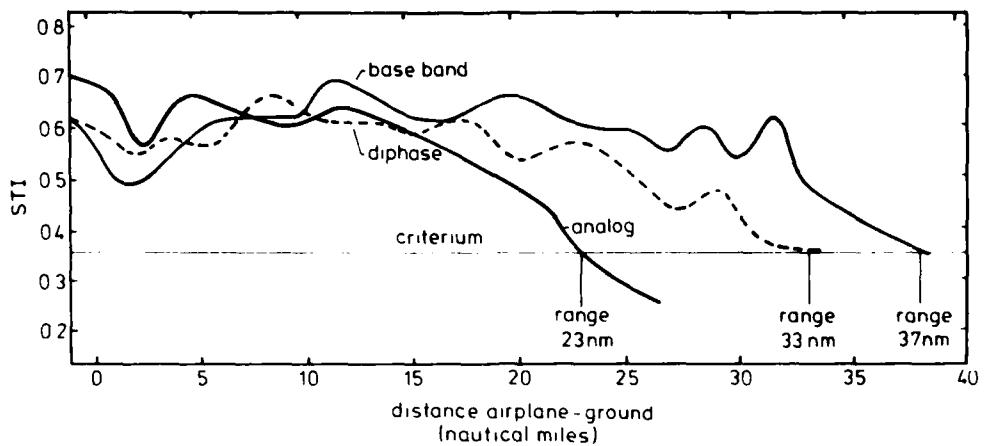
Fig. 6   Example of STI as a function of range between an airplane and a
ground station for an analog transmission link and a CVSD system.

## 3. EVALUATION OF A DIGITAL COMMUNICATION SYSTEM

In a comparative investigation the performance of two types of secured voice communication systems both based on Continuous Variable Slope Deltamodulation (CVSD) was studied. From the two systems the intelligibility on basis of STI was measured as a function of the bit error rate between coder and decoder.

In Fig. 5 the STI as function of this bit error rate is given for system A at two bit rates (16 Kbit/sec and 8 Kbit/sec) and for system B at 16 Kbit/sec only. These results show a better performance of the CVSD-A 16 Kbit/sec system. From this it was concluded to compare the performance of the CVSD-A 16 Kbit/sec system with an analog system on a VHF communication link between an airplane and a ground station. At the transmission side in the plane the test signal was obtained from a taperecorder with the prerecorded test signal. At the ground station a real-time analysis of the received test signal was performed. In this way the STI as a function of the distance between airplane and ground station was obtained. Fig. 6 gives the results for 3 transmission methods:

a. CVSD-A baseband modulation
b. CVSD-A diphase modulation
c. Analog channel

As criterium for the maximum range a STI of 0.35 is used. This means a range of respectively 37, 33 and 23 nm for the 3 transmission methods.

Fig. 7   STI as a function of the background noise for the five investigated microphones at a distance of 0 cm.
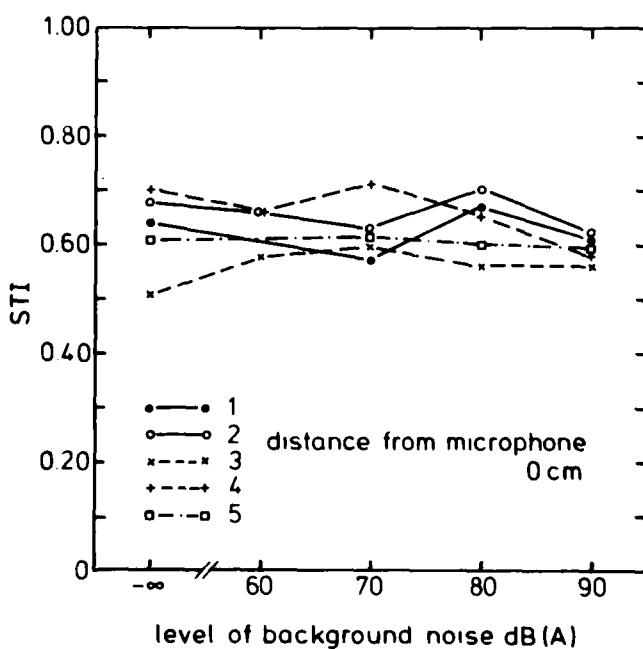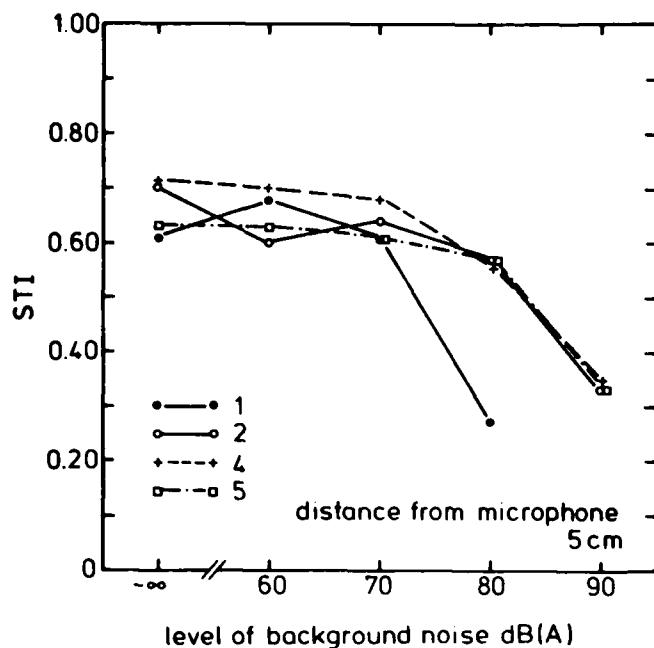
Fig. 8 STI as a function of the background noise for four of the microphones at a distance of 5 cm.

4. EVALUATION OF MICROPHONES AND TELEPHONES

In a comparative investigation the intelligibility of five sound powered microphone/ telephone systems was studied. As the system had to be used in a noisy environment the intelligibility was measured as a function of the level of background noise. The intelligibility was measured with the STI-method. At the transmission side a sound source mounted in an artificial head was used. The microphone under test was placed in front of the mouth from which the test signal originated. The output signal of the microphone was connected to the STI analyzer. The STI of this measuring setup was measured as a function of the level of acoustically added background noise and for 2 distances of the microphone from the mouth. The level of the test signal in front of the mouth was adjusted to the level of normal speech. Figs. 7 and 8 give the results of these measurements for the five microphones used in the experiment. Although the differences between the microphones are relatively small, the best results are obtained with microphone no. 4 for both conditions. The same type of experiment was performed with the telephones of the systems. Here the test signal was connected electrically to the telephone cartridge. The telephone was placed on the head of a subject. The test signal coming from the telephone was picked up by a small electret microphone (8 x 6 x 2 mm, placed near the ear entrance of the subject) and con-

Fig. 9 STI as a function of background noise for the five investigated telephones.
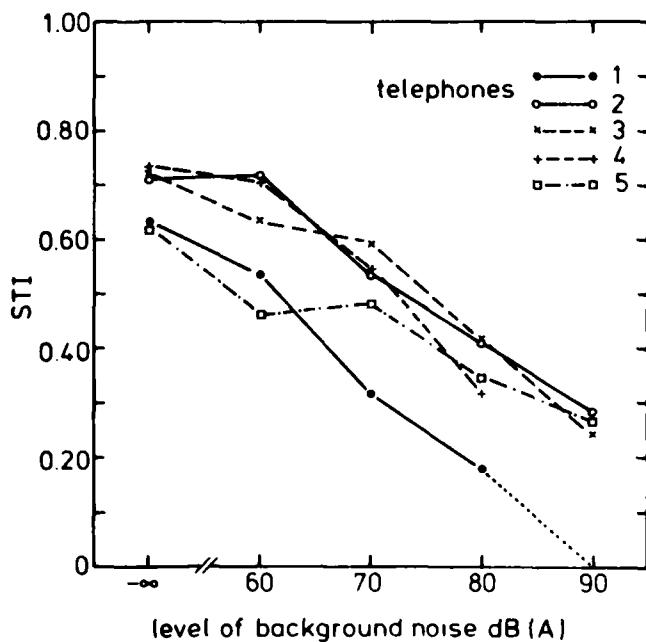
nected to the STI analyzer. The STI, measured as a function of the background noise level
is given for the five investigated systems in Fig. 9.

5. CONCLUSIONS

The relation between STI and subjective intelligibility obtained on digital channels
offers the same accuracy as was found before on analog channels and in room acoustics.
Because of the relatively short measuring time a number of measurements can be performed
which were not possible with the traditional subjective methods.

REFERENCES

(1) Houtgast, T., and Steeneken, H.J.M. (these proceedings, 1981). STI – an Objective
    Measure for the Performance of Voice Communication Systems.

(2) Steeneken, H.J.M., and Houtgast, T. A Physical Method for Measuring Speech Transmis-
    sion Quality. J. Acoust. Soc. Amer. 67 (1980) 318-326.

(3) Steeneken, H.J.M., and Agterhuis, E. Institute for Perception TNO, Soesterberg, The
    Netherlands. Description of STIDAS II-C (Speech Transmission Index Device using
    Artificial Signals) Part I (1978) Report no. IZF 1978-19.

DISCUSSION

DR. J.D. MOSKO (UNITED STATES)

Have you had an occasion to use the STI for the placement of speaker systems in work spaces
onboard ships?

DR. H.J.M. STEENEKEN (NETHERLANDS)

Yes, we studied the gradient of the STI in rooms as a function of place, and are currently
involved in a study where the intelligibility in command rooms and cross talk between different
operators is a point of interest.

VOICE COMMUNICATION CAPABILITY OF SELECTED INFLIGHT HEADGEAR DEVICES

Richard L. McKinley
Charles W. Nixon
Thomas J. Moore
Air Force Aerospace Medical Research Laboratory
Wright-Patterson Air Force Base, Ohio 45433

SUMMARY

The voice communications effectiveness (MRT word intelligibility) of
selected Air Force communications terminal equipment was evaluated in
simulated operational noise environments. Analyses of the resulting
data indicate:

1. Standard AF communications headsets H-133, HGU-26/P and H-157
performed in a manner consistent with their design purposes with the
H-133 providing the best communication, the HGU-26/P second and the H-157
third.
2. Percent correct intelligibility in the 115 dB noise condition was
reduced as much as 15% for the H-133 and 50% for the H-157 over the ambient
noise condition.
3. Communication performance varied over 5 to 10 percent when used with
the AIC vs the AIC and RF radio.
4. The UK chemical defense hood provided a slightly better talking
environment than the MBU-5/P (~3%) and a worse listening environment than
the HGU-26/P (4-6%).
5. The new "thin" M-101 microphone provided better intelligibility (6%)
than the standard M-101 microphone.

INTRODUCTION

Effective voice communication is a critical requirement for the successful execution of military
operations. Continuing efforts to improve the effectiveness of these operations and activities typically
require the development of new, advanced and more complex personnel equipment and systems. Among the
systems and equipments that are current or are intended for future use are several items that mandate
efficient voice communications, especially in the presence of high level acoustic environments. However,
many of these personal systems and equipment items have features that reduce their noise excluding properties
and allow an increase in interference with audio communications. In addition, many of the noise environ-
ments in which these units must be worn have increased in level and duration of exposure. In view of these
factors, the audio communications effectiveness of selected inflight headwear units were evaluated in
specified noise environments.

Current USAF terminal voice communication equipment has been designed to effectively operate in the
typical noise fields in which they are utilized. However, other items have been designed with voice
communication and noise exclusion performance secondary to different requirements for personnel safety.
New devices are being developed primarily for chemical and biological protection in addition to voice
communication capabilities. Advances in military aircraft performance have generated new requirements
for lighter and more stable headgear for acceptable use in the High G-load environments of advanced
weapons systems. During the development of new systems such as these interim prototypes and models must
frequently be modified and redesigned to insure adequate voice communications capability in the final
product. Investigations of the current and new systems must be carried out under simulated conditions
that represent the operational environment with emphasis on the acoustic environments because of their
persistent threat to effective voice communications

PURPOSE

The purpose of this report is to describe the voice communications effectiveness of selected items of
personal equipment used in the presence of simulated operational noise environments. The USAF standard
inflight and ground voice communication equipment, a United Kingdom chemical defense hood, and a small
M-101 microphone (see Table 1) were evaluated and the results are presented herein.

TABLE 1
UNITS EVALUATED FOR VOICE COMMUNICATION CAPABILITY

USAF HGU-26/P Flight Helmet with MBU-5/P Oxygen Mask and M-101 Microphone	(AF Standard)
USAF H-157 Inflight Headset with M-87 Boom Microphone	(AF Standard)
USAF H-133 Ground Communications Headset with Microphone Noise Shield and M-101 Microphone	(AF Standard)
United Kingdom Chemical Defense Hood	
M-101 Microphone Element (smaller size than the standard unit)	

METHOD

Approach

The comparative intelligibility of standardized test materials processed through the representative communication headgear listed in Table 1 was measured in the presence of varying levels of simulated operational noise. Volunteers spoke and responded to the speech communication signals under the specified experimental conditions. Variations in measured intelligibility were attributed to the voice communication capabilities of the different units and to the interference effects of the noise conditions.

Subjects

Ten subjects, five male and five female, were employed in the investigations. The subjects were recruited from the general civilian population. Each subject participated in at least eight hours of training before data collection began. The subjects were paid an hourly rate for participation in the experiments and a bonus payment was made to all subjects who completed all test conditions. Subjects were otologically normal and had hearing levels no greater than 15 dB at any standard audiometric test frequency from 125 to 8000 Hz.

Facilities

Communication Stations. These studies were accomplished utilizing the Voice Communication Research and Evaluation System (VOCRES) of the United States Air Force, Aerospace Medical Research Laboratory. This research facility has the capability to realistically model the major acoustic factors experienced by crew members that may adversely affect voice communications. The overall system includes a master control station and ten individual aircraft communication stations. Each station contains the USAF standard intercommunication system (AIC/25) and respiration system (A-19). Both intercommunication and respiration terminals and operating controls are easily accessible to the individual positioned at the station. Each station is also integrated into a Computer-Display-Response system in which the central processor is a Hewlett-Packard 9845T. An interface at each station decodes commands by the central processor to the station's display and also returns the subject's response to the central processor for storage and analysis.

Aircraft Radios. The transmitters and receivers used in the study were ARC-150 radios, a current operational US Air Force UHF aircraft radio. The communication signals were transmitted and received by the radios and then presented to the listeners through the standard Air Force aircraft intercommunication system. All connections between the RF transmitter and receiver were made by means of standard 50 ohm coaxial cable.

Acoustic Environment Simulations. The acoustic environment simulation facility consists of a large reverberation chamber (approximately 8000 ft^3) that houses a powerful electrodynamic sound system. The electrodynamic system contains dual amplifiers that may be used singly or in combination. One system (low power) consists of two 600-watt amplifiers and the other (high power) consists of two 7000-watt amplifiers. The amplifiers drive 8 loudspeaker banks, each containing twelve 15-inch loudspeakers, forty-eight 3-inch high frequency "tweeters" and one midrange compression driver. The loudspeaker enclosures are portable and may be rearranged for various purposes. In the configuration used for these studies, the low power system generated a maximum overall Sound Pressure Level (SPL) of 122 dB, while the high power system generated a maximum overall SPL of 128 dB re 20 µPa (with a pink noise input). The spectra and levels of the noise conditions used in this study are presented in Table 2.

The low power system was used in the cockpit noise environment simulation. A pink noise source was shaped by a 1/3 octave band spectrum shaper so that the spectrum measured in the test space was representative of that experienced in the cockpit of a typical jet fighter aircraft.

TABLE 2

SPECTRA AND LEVELS (dB SPL) OF NOISE CONDITIONS USED IN
EVALUATION OF COMMUNICATIONS EQUIPMENT

OCTAVE BAND CENTER FREQUENCY (Hz)	95 dB	105 dB	115 dB
31.5	76	87	96
63	90	100	109
125	88	98	109
250	85	96	106
500	81	92	102
1000	81	92	102
2000	88	99	109
4000	84	96	105
8000	76	87	97

Measurement Instrument

This study employed a standardized measure of intelligibility, the Modified Rhyme Test (MRT) as developed by House, et al (1963), for assessing communication effectiveness. The MRT was selected for use over other test materials because of evidence that it is the test of choice for evaluating the performance of military speech communication systems in the presence of environmental noise (Webster and Allen, 1972). The materials consist of lists of 50 one-syllable words that are equivalent (lists) in intelligibility. These test words are presented embedded within a carrier phase that is the same for each item. The MRT is easy to administer, score and evaluate and it does not require extensive training of listeners. The display at the talker's station (Fig 1) provided the text which the talker read, while the listeners' stations displays (Fig 2) provided a choice of six key words from which the listener

selected the one he/she felt was correct by pressing the button beside it. To compensate for correct
answers obtained by guessing a correction factor was applied to the scores.
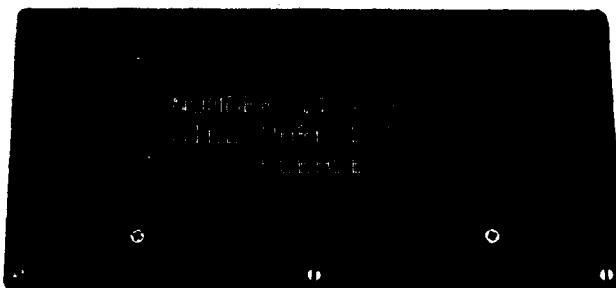

FIGURE 1
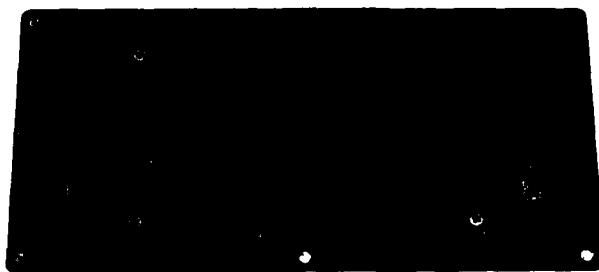
FIGURE 2

Experimental Procedure

The fundamental procedure utilized in AFAMRL voice communication studies involves Round Robin
type paradigms wherein selected subjects perform as talkers and all subjects participate as listeners.
The talker on any one trial serves as a listener on previous and subsequent trials while non-talkers per-
form as listeners on all trials. The number of talkers and listeners for an evalaution and consequently,
the test design employed, is determined by the number of communication units available for the test.
Generally, five talkers, three male and two female are selected from the subject panel of ten members.
Each trial consumes about 5 1/2 minutes. A daily test session covers a period of about four hours with 4J
minute test sessions separated by 15 minute rest breaks. All ten subjects participate in the experiment
simultaneously, as stated earlier, with one acting as the talker and the other nine as listeners.

The three standard USAF communication units were evaluated in accordance with the general procedures
just described at four different noise levels. These headset-microphone combinations were evaluated with
two different experimental arrangements. The first configuration, as indicated in Figure 3, involved only
the terminal equipment and the standard aircraft intercommunication system (AIC). The second configuration,
as shown in Figure 4, included an RF link, such that the output of the talkers AIC was transmitted by one
ARC-150 and received by another before presentation to the listeners via the AIC and listener terminal
equipment. All subjects participated in all sessions, essentially serving as their own controls.
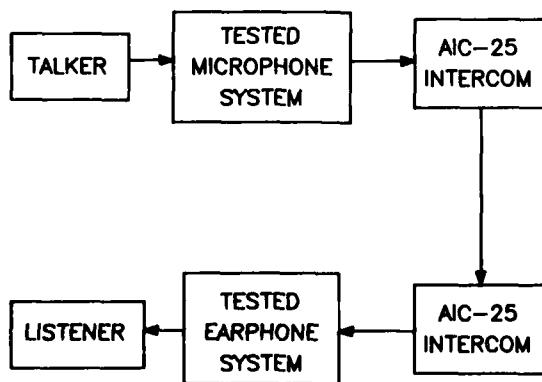

FIGURE 3
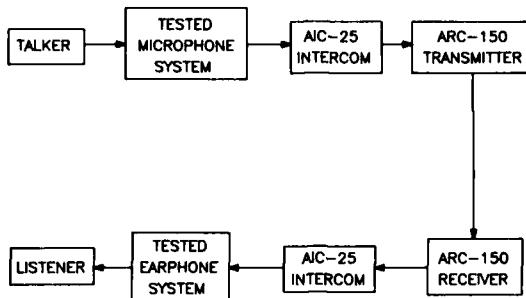
FIGURE 4

Only three models of the UK Chemical Defense hood were available, consequently the general test design was modified for the evaluation. Three talkers were selected from the subject population, two male and one female, for the measurements. Each of the three talkers spoke with both the UK chemical defense hood and with the MBU-5/P oxygen mask two times each, thereby serving as their own controls. Listeners wore the chemical defense ensemble and the HGU-26/P whose performances were compared. Measurements were made in an ambient noise level of 105 dB.

The microphone element investigation used a complete round robin design wherein each of the ten subjects participated as talker. This procedure was employed to collect data on both the standard microphone element and the test microphone element that was designed to be physically smaller than the standard while providing equal or better performance. The microphones were evaluated in the single ambient noise at an overall level of 105 dB.

RESULTS AND DISCUSSION

The data collected on the various communications units using the MRT were tabulated in terms of percent correct intelligibility scores and treated with measures of central tendency and variance statistics.

Air Force Standard Units. The average performance of the three AF communication units in the ambient noise conditions are summarized in Figure 5, which used the aircraft intercommunication system only and in Figure 6 which used both the AIC and the UHFARC-150 aircraft radio. Overall, the comparative intelligibility between units and as affected by the noise are consistent with the design goals of the units. The H-133 unit with microphone shield was designed for use in noise environs that reach and exceed 135 dB SPL. The HGU-26/P helmet and oxygen mask systems were designed for use in cockpit noise environs that reach and exceed 115 dB SPL. The H-157 with boom microphone was intended for moderate noise environs that may reach intensity levels of around 95 dB SPL. Overall, the H-133 provided the best communication of the three followed in order by the HGU-26/P and the H-157, except for AIC only data in Figure 3 where the H-133 and HGU-26/P were essentially the same.


FIGURE 5

Intercom with ARC-150

H-133   HGU-26/P   H-157

% Correct (Adjusted for Chance)
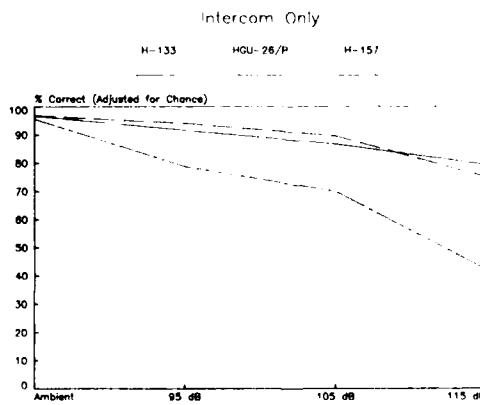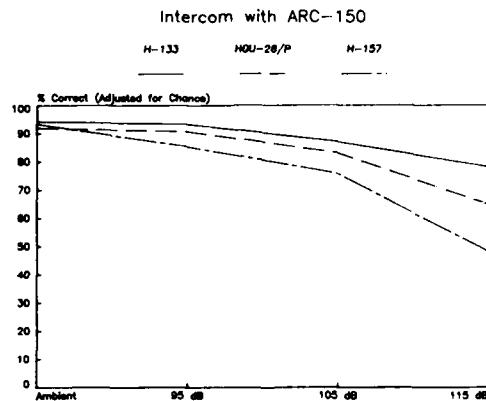
FIGURE 6

As the levels of the noise conditions were increased above the ambient condition the communication performance of the units decreased, both for the AIC and the AIC plus RF conditions. Under the ambient condition (about 79 dB SPL) all units were found to be equally effective with percent correct intelligibility scores of about 92 to 96 percent. The performance of the H-133 declined with increased noise level but remained acceptable at the 115 dB condition. If the slope of the performance line remains constant a percent correct intelligibility would be estimated at 70 to 75 percent at 125 dB and 65 to 70% at 135 dB for the noise spectrum used. In noise spectra found during many ground maintenance activities the greater low frequency content of the acoustic energy could result in higher voice communication scores than estimated from these data.

The communications effectiveness of the HGU-26/P helmet decreases more rapidly than the H-133 with increasing noise, however, it does remain adequate (around 70% correct) in the 115 dB SPL condition. Extrapolations to noise levels of 125 dB and 135 dB estimate the helmet system to be marginal or inadequate with percent correct estimates of 45 to 60% and 30 to 45% respectively. The H-157 headset system provided satisfactory communications in the 105 dB SPL noise condition but was not acceptable in the 115 dB SPL noise condition, as predicted. These measurement results confirm that the three standard USAF headset-communication systems perform satisfactorily in the noise environments for which they were designed. Although the basic design specifications were developed over 20 years ago, the soundness of the research developed technology base upon which the design specifications were formulated is clearly evident.

The MRT average intelligibility scores for the USAF units are shown with the AIC only and with the AIC and aircraft radio in Figures 7, 8 and 9. It is clear from Figure 7 that communications effectiveness at these noise conditions is unaffected by the addition of the aircraft radio, with satisfactory communications for all conditions. The HGU-26/P (Figure 8) shows a decrease in percent correct intelligibility with the radio link added of about 5% under ambient and the 95 dB SPL noise condition and increasing to about 10% at the 105 dB and 115 dB SPL noise conditions. Operational voice communications should remain satisfactory for all these conditions. The communication performance of the H-157 (Figure 9) is essentially equivalent under ambient conditions but is about 5% better under the three noise conditions with the RF radio in the communications link. Although voice communications may be considered adequate with this unit in 95 dB and 105 dB noise conditions it is considered unacceptable at the 115 dB SPL condition.

H-133

AIC Only   AIC with Radio

% Correct (Adjusted for Chance)

FIGURE 7

HGU-26/P

AIC Only   AIC with
                  Radio

% Correct (Adjusted for Chance)

FIGURE 8

H-157

AIC Only   AIC with
                  Radio

% Correct (Adjusted for Chance)
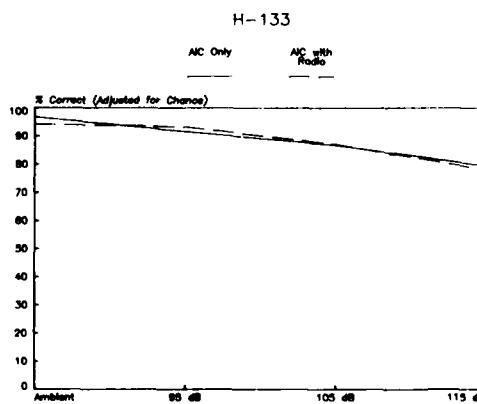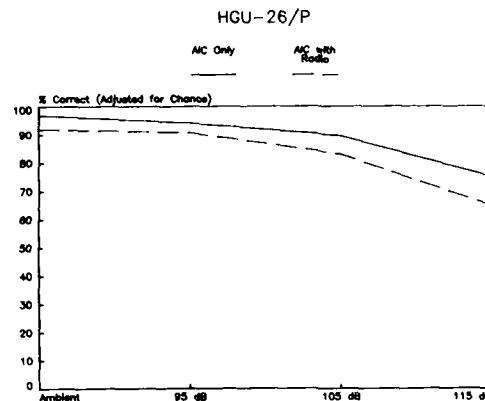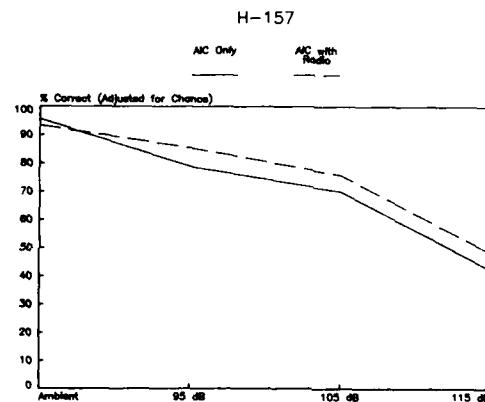

FIGURE 9

Data presented on terminal voice communications equipment using the aircraft intercommunication system with and without the aircraft radio emphasizes the importance of the total system in a laboratory evaluation scheme. Some of the differences reported here are of sufficient magnitude to influence results. Laboratory measurements which use "high fidelity" or other kinds of band pass schemes in their test setups will be influencing the performance attributed to the terminal equipment with these schemes. It is repeated here, that the laboratory measurement system should duplicate/simulate the system and environmental variables in which the test item is to be used to the fullest extent that is practical, to provide as much validity as possible and maintain laboratory credibility with operational personnel and activities.

United Kingdom Chemical Defense Hood. A comparative evaluation of various combinations of components of the UK chemical defense hood and the HGU-26/P helmet system were accomplished in a single noise environment of 105 dB SPL with the AIC and RF radio configuration of Figure 2. The results of these measurements are contained in Table 3 in the form of mean percent correct intelligibility and standard deviation scores. The various combinations of systems were arranged to allow overall comparisons of the UK Chemical Defense Hood vs the US HGU-26/P systems and estimations of the talking vs the listening capabilities of the sub-units. The mean percent correct intelligibility score for the HGU-26/P (listeners) with MBU-5/P (talkers) system was 86% which is essentially identical to that measured earlier and reported on Figure 6 for the 105 dB SPL noise environment. The mean intelligibility score increased to 89% using the UK chemical defense hood for the talkers and the HGU-26/P helmet for the listeners. When the listeners wore the UK unit the mean intelligibility scores were about 82% whether the talkers used the same UK unit or the MBU-5/P oxygen mask microphone. These data are interpreted to suggest that the UK chemical defense hood provided a slightly better talking environment than the MBU-5/P (~3%) while providing a worse listening environment than the HGU-26/P (4-6%). Generally differences among these 4 combinations are small (2% - 6%) and the units would be expected to provide about the same degree of communications effectiveness in the operational environment. The scores of 82% to 89% would indicate satisfactory communications for the spectra and levels of 115 dB SPL as utilized in this study.

TABLE 3

COMPARATIVE PERFORMANCE OF THE UK CHEMICAL DEFENSE HOOD
AND THE USAF HGU-26/P INFLIGHT HELMET SYSTEM

EXPERIMENTAL CONDITION	MEAN WORD INTELLIGIBILITY	STANDARD DEVIATION	N
Talker: MBU-5/P Listener: HGU-26/P	86	7.5	24
Talker: UK CD Hood Listener: HGU-26/P	89	7.4	30
Talker: MBU-5/P Listener: HGU-39/P & UK CD Hood	82	8.3	18
Talker: UK CD Hood Listener: UK CD Hood & HGU-39/P	83	8.5	12

Thin M-101 Microphone Element. The word intelligibility of the thin M-101 microphone element was compared to that of the AF standard M-101 microphone in the 105 dB SPL noise condition. Results of these measurements are summarized in Table 4.

TABLE 4

PERCENT CORRECT INTELLIGIBILITY (MRT) OF STANDARD
VS "THIN" M-101 MICROPHONE ELEMENT

SUBJECT	"THIN" M-101 MICROPHONE ELEMENT	STANDARD M-101 MICROPHONE ELEMENT
1 M	87	73
2 M	77	75
3 F	82	81
4 M	93	83
5 M	87	76
6 F	87	83
7 F	63	63
8 M	80	71
9 F	89	83
10 F	84	80
MEAN PERCENT CORRECT (MRT)	83%	77%

The new "thin" microphone design provided about 6% better performance than the standard microphone when evaluated by the ten subjects (five male and five female). Inspection of the data reveals that nine of the ten subjects obtained better performance with the new microphone element design while none obtained better performance with the standard unit. The range of improvement was 0 to 11% over all subjects. The new microphone should be evaluated across a range of noise spectra and levels to fully define its performance profile.

SUMMARY

The voice communications effectiveness (MRT word intelligibility) of selected Air Force communications terminal equipment was evaluated in simulated operational noise environments. Analyses of the resulting data indicate:

1. Standard AF communications headsets H-133, HGU-26/P and H-157 performed in a manner consistent with their design purposes with the H-133 providing the best communication, the HGU-26/P second and the H-157 third.
2. Percent correct intelligibility in the 115 dB noise condition was reduced as much as 15% for the H-133 and 50% for the H-157 over the ambient noise condition.
3. Communication performance varied over 5 to 10 percent when used with the AIC vs the AIC and RF radio.
4. The UK chemical defense hood provided a slightly better talking environment than the MBU-5/P (~3%) and a worse listening environment than the HGU-26/P (4-6%).
5. The new "thin" M-101 microphone provided better intelligibility (6%) than the standard M-101 microphone.

DISCUSSION

MR. G. GARINTHER (UNITED STATES)

Why do you use MRT test rather than PB testing?

To what criterion level do you train your subjects?

MR. R.L. MCKINLEY (UNITED STATES)

PB testing requires extensive training, while the MRT does not. A draft ISO Standard designates the MRT as the test of choice for speech intelligibility testing.

Subjects are trained to 90% correct MRT words using helmets and oxygen masks.

DR. H.M. BORCHEGREVINK (NORWAY)

What signal-to-noise ratios did you use?

Were the subjects allowed to adjust the volume of speech signal?

Was the speech SPL adjusted so high that it might cause hearing damage under the signal-to-noise ratios used in your experiments?

MR. R.L. MCKINLEY (UNITED STATES)

If you would contact us by letter we would be glad to provide the signal-to-noise ratios.

Yes, during the training period the subjects were allowed to adjust the volume control. During the experimental period the subjects were instructed not to change the volume control.

The average preferred listening level was 95 dB. During long term flights this is sufficient to present a risk to hearing. However, during the time duration of the experiment this level was not hazardous to hearing.

DR. W. ENDRICH (NETHERLANDS)

At the start of your presentation, you said that you would have liked to have the system described by Dr. Houtgast in the previous paper (STI - an Objective Measure for the Performance of Voice Communication systems); we have heard that the results with this system become quickly available (1-2 minutes). Pragmatic question:

i) How long did your analysis take?

Further question:

ii) Is there a possibility for quickly relating the results of your studies to an absolute standard of comparison, such as STI?

MR. R.L. MCKINLEY (UNITED STATES)

Two weeks.

If Dr. Houtgast's system were available in the U.S., yes, correlation would be possible.

By backtracking through the articulation index, it might also be possible to relate our results to the STI.

AN AUTOMATED MULTIPLE CHOICE INTELLIGIBILITY TESTING SYSTEM

by

R.L. Pratt

Royal Air Force Institute of Aviation Medicine

Farnborough Hampshire England GU14 6SZ

INTRODUCTION

The Royal Air Force Institute of Aviation Medicine (RAF IAM) conducts intelligibility tests as part of a programme to evaluate the quality of various items which may affect aircrew communications (e.g. microphones). In aircraft where an oxygen mask is mandatory, the intelligibility of the speech signal (when recorded direct from the mask microphone) is generally satisfactory for two reasons:

a. The close fit of the mask to the face serves to reduce the cockpit noise level picked up by the microphone.

b. By enclosing the mouth in this way the sound pressure level (SPL) of the speech is raised.

The combination of these two factors thus gives rise to a condition of high signal to noise ratio (SNR).

In aircraft that do not employ an oxygen system, but do suffer from high cockpit noise levels (particularly rotary wing aircraft) the situation is somewhat different. The traditional solution is to use a throat microphone, which derives its signal from the mechanical contact with the throat of the wearer and has the advantage of low sensitivity to cockpit noise. The disadvantage with this method is the fact that the signal generated by the displacement of the throat lacks frequency components generated by resonances in the vocal tract, and by the lips, which are of critical importance in the generation of intelligible speech. An alternative to the throat microphone is a 'first order pressure differential microphone' (sometimes referred to as a noise cancelling microphone), mounted on a boom. Such devices have been issued to aircrew when, for medical reasons, a throat microphone is not acceptable. Early microphones of this type tended to have poor noise cancelling properties which resulted in a high degree of pickup of cabin noise. A new generation of microphones is now available with a considerable improvement in this respect, and it is these microphones which are currently being investigated at the IAM.

As a result of the improved speech quality of such microphones, intelligibility experiments conducted in simulated rotary wing cabin noise can give results where the percentage of correct word identification exceeds 90%. This fact combined with the similarity of performance of the microphones under consideration results in a situation where the percentage of words correctly identified may not be a sufficiently sensitive measure to enable fine differences in performance to be established.

One of the methods to improve the sensitivity of intelligibility tests discussed in a paper by Hecker et al. (1966) involves the measurement of a subject's response time to test words. In these experiments five subjects listened to six Modified Rhyme Test (MRT) lists in six signal to noise ratios (30, 20, 15, 10, 5, 0 dB) and their responses, together with the test word were fed to a 2 channel pen recorder, thus enabling a measure of reaction time to be made. It was found that as the SNR was reduced the reaction time increased, and the percentage of correct identifications fell. The authors therefore claimed an increase in sensitivity for the method.

EXPERIMENTS AND RESULTS

The technique developed at the Institute employs a digital computer which not only administers the test automatically, but also checks the subject's response for accuracy and computes the reaction time.

The test material used is the Modified Rhyme Test and the Clarke's Vowel Test (CVT). The MRT consists of groups of single syllable words which differ only in their initial or final consonant; for the CVT only the vowel is different. In the experiment reported here the relative performance of various noise cancelling microphones was compared by recording MRT and CVT words lists against a background of rotary wing noise provided using a pre-recorded sample formed into a tape loop. The word lists were then re-played to subjects, seated in the same noise environment used for the recordings, through a communications headset with attenuation properties similar to helmets used in rotary wing aircraft. Subjects were required to select the word they thought they heard from a group of six words (five in the case of the CVT) presented on a Visual Display Unit (VDU).

The system is based on a PDP 11/03 and a block diagram is given in figure 1. An adjustable attenuator is included in the replay chain to permit subjects to choose their own preferred listening levels. The word list recordings are also fed to a threshold detector (Bruel and Kjaer Narrow Band Analyser Type 2031) which generates an interrupt to the computer when the beginning of a word is detected. The computer then displays a multiple choice selection on a VDU and a subject responds by pressing one of a series of buttons mounted on the screen alongside the displayed words. The computer is thus able to determine the accuracy of the response, and in addition it determines the subject's response time to a word.

In order to investigate the characteristics of this system further results obtained from it were compared with those obtained using a conventional (or "manual") system in which the test subject identified the word from a printed list. Two groups each of six subjects participated in this experiment. One group completed a Modified Rhyme Test using the automatic system and a Clarke's Vowel Test manually, followed approximately one week later by a Modified Rhyme Test administered manually and a Clarke's Vowel Test automatically. For the other group the order of these sessions was reversed.

Results were analysed using analysis of variance and it was found that, for the MRT, there was no statistically significant difference between the mean scores for the automatic and manual test. Furthermore, there was no difference between the mean scores for the two sessions, indicating that there is a minimal learning effect. With the CVT the automatic scores were significantly lower ($p < 0.01$), but the difference itself was small (5.4%). These results suggest that the automatic administration of the tests does not have a major influence on subjects' performance.

An experiment has been conducted in which a reference noise cancelling boom microphone was compared with three test microphones using both MRT and CVT, during a total of six paired comparisons. Using the student's t test the percentage of correctly identified words enabled only one discrimination between a pair of microphones to be made. However, using the response times as a measure of intelligibility enabled three discriminations to be made. Thus, it would appear that a consideration of the subject's reaction time may increase the sensitivity of the test and permit finer distinctions to be made.

CONCLUSIONS

An automated intelligibility testing system has been constructed and was found to give results very close to that of the conventionally administered test. It is suggested that subjects' reaction time, when used in conjunction with the percentage correct score, may assist in discriminating between microphones of comparable performance.

REFERENCE

Hecker,M.H.L., Stevens,K.N. and Williams,C.E. (1966).   Measurement of Reaction Time in Intelligibility
   Test.   Journal of the Acoustical Society of America 39 p.1188.
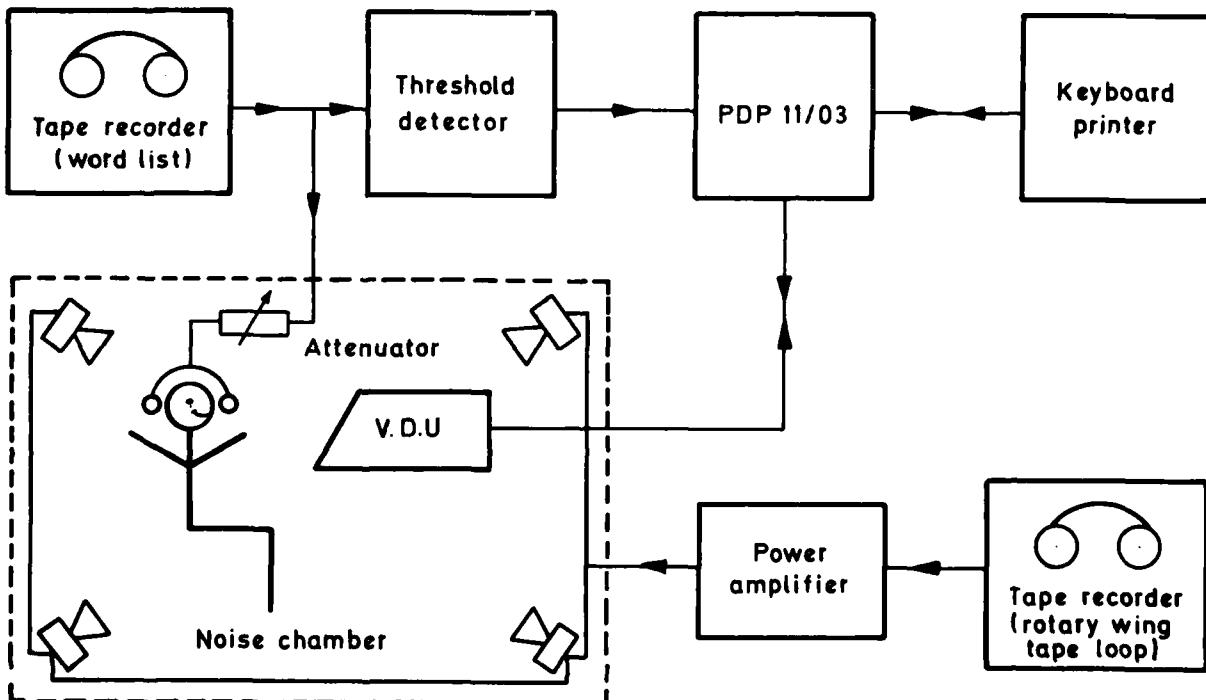
Fig 1 Block diagram

DISCUSSION

DR. J.D. MOSKO (UNITED STATES)

Have you had the opportunity to look at RT's to word initial and word final positions?

Have you had the opportunity to look at the PBT developed by Mitchell a few years ago at CCNY?

What was the relation between the response time window for the automated test and the "write-down" version?

DR. R.L. PRATT (UK)

The answer to questions 1 and 2 is NO, because the system is intended to assist the selection of aircrew transducers rather than a speech intelligibility research tool.

DR. R.L. MCKINLEY (UNITED STATES)

Do you think that a change in your computer program to present the visual display of the words before the auditory stimulus would reduce the difference between the automated and manual tasks?

Did your subjects receive equal training time on both the manual and automated tasks?

DR. R.L. PRATT (UK)

I think it probably would, but the difference between the two systems is so slight that it is of little practical importance.

Yes. All subjects completed several word lists with each system under conditions of high SNR (approximately 30 dB) and always scored in excess of 95% words correct.

EFFECTS OF AGE, FLYING TIME AND TYPE OF AIRCRAFT ON THE HEARING OF GERMAN MILITARY PILOTS, AND IT'S SIGNIFICANCE FOR INFLIGHT COMMUNICATION

by

G.R. Fröhlich, Col., GAF, MC
German Air Force Institute of Aviation Medicine
P.O. Box 172 KFL, D-8080 Fürstenfeldbruck, Germany

INTRODUCTION

Since 1979 our Hewlett & Packard Computer has been operational and all pilot audiometric data from 125 - 8000 Hz have been stored. This made it possible to compute the means and standard deviations of the pure tone audiograms of 4,034 pilots investigated in 1979.

METHODS

These pilots have been divided into different categories:
a) Pilots of jet aircraft (J), propeller and turboprop aircraft (P) and helicopters (H)
b) Age: 20 - 30, 31 - 40, above 40 years
c) Flying hours: 1 - 1000, 1001 - 3000, above 3000 hours
The audiograms have been subdivided into:
a) normal hearing: no more than 20 dB hearing loss at 1000 - 6000 Hz
b) slight sensory-neural hearing losses within standards
c) marked hearing losses requiring a waiver
d) all other forms of deafness have not been evaluated, since their total incidence was below 1%.

German Audiometric Standards:

Candidates: no more than 20 dB losses at 250 - 2000 Hz; at 3000, 4000 and 6000 Hz the hearing loss must not exceed a total of 210 dB for both ears. For pilots of all types of aircraft:no more than 30 dB of hearing loss at 250 - 2000 Hz.

RESULTS

In age groups 20 - 30 years and/or 1 - 1000 hours ($\bar{x}$=400), (Fig. 1 and 2) more than 90% of the pilots revealed normal hearing, the rest had slight hearing losses as shown in the example for jet pilots:

		1000	2000	3000	4000	6000 Hz
r	9%	$10^{+}_{-}6$	$9^{+}_{-}5$	$15^{+}_{-}10$	$27^{+}_{-}16$	$36^{+}_{-}17$ dB
l	10%	$9^{+}_{-}4$	$9^{+}_{-}6$	$16^{+}_{-}11$	$28^{+}_{-}17$	$38^{+}_{-}16$ dB

There was no waiver necessary.

%Distribution of Hearing Losses Versus Flying Hours

J = Jet
P = piston-engined + Turboprop
H = helicopter
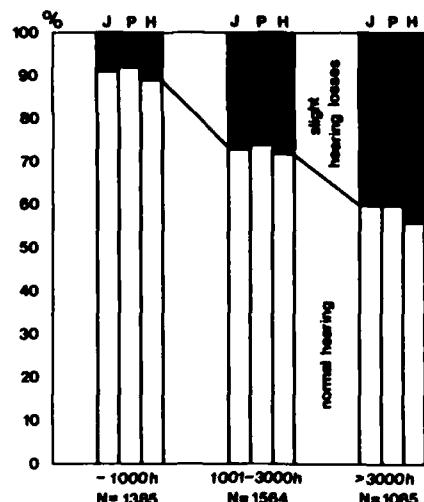
Figure 1

% Distribution of Hearing Losses Versus Age

J = Jet
P = piston-engined + Turboprop
H = helicopter
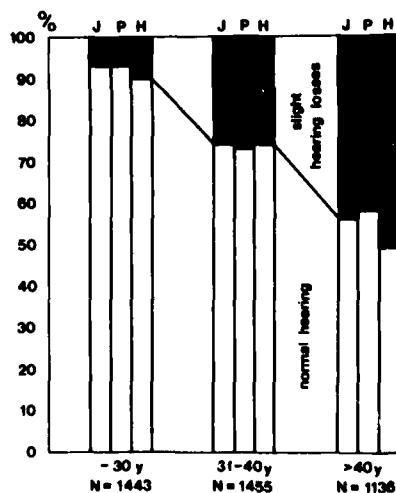
Figure 2

In age groups 31 - 4o years and/or 1oo1 - 3ooo flying hours ($\bar{x}$=1985), 7o - 75% of all pilots revealed normal hearing. The rest showed slight hearing losses as indicated by the example for helicopter pilots:

		1ooo	2ooo	3ooo	4ooo	6ooo Hz
r	25%	1o$\pm$4	11$\pm$6	21$\pm$15	37$\pm$17	39$\pm$18 dB
l	3o%	9$\pm$4	11$\pm$6	22$\pm$14	36$\pm$18	4o$\pm$19 dB

1% of helicopter pilots showed marked hearing losses on either right or left ears requiring a waiver. Otherwise there were no differences between pilots of different types of aircraft.

In age groups above 4o years and/or above 3ooo hours flying time ($\bar{x}$=4ooo), still more than 5o% of all pilots revealed normal hearing. The incidence of slight hearing losses well within standards was in jet pilots 36% of right and 43% of left ears; in helicopter pilots 4o% of right and 42% of left ears as shown in the example below:

		1ooo	2ooo	3ooo	4ooo	6ooo Hz
r	4o%	1o$\pm$5	11$\pm$6	19$\pm$13	35$\pm$17	36$\pm$17 dB
l	42%	9$\pm$4	12$\pm$7	22$\pm$14	38$\pm$17	4o$\pm$16 dB

The incidence of marked hearing losses in jet pilots was 2% for right and 3% for left ears, whereas for helicopter pilots it was 3% for right and 4% for left ears as shown below.

		1ooo	2ooo	3ooo	4ooo	6ooo Hz
r	3%	18$\pm$12	28$\pm$17	52$\pm$25	58$\pm$24	56$\pm$24 dB
l	4%	23$\pm$12	33$\pm$19	53$\pm$23	58$\pm$23	53$\pm$19 dB

The total of all marked hearing losses comprised 2o right and 32 left ears, that is o.65% of all ears. These pilots were subjected to speech audiometry (short sentences) in 8o dB white noise and revealed 1oo% intelligibility in higher speech levels, so that waivers could be granted.

Table 1

Audiograms of Jet Pilots Versus Flying Hours

Flying hours	Category of hearing losses	%		Hearing Losses (dB)				
			1000	2000	3000	4000	6000 Hz
1 - 1000	normal	91 r	9$\pm$3	8$\pm$4	8$\pm$4	10$\pm$5	12$\pm$6
$\bar{x}$ = 315		90 l	7$\pm$4	7$\pm$4	8$\pm$5	10$\pm$5	11$\pm$6
N = 801	slight	9 r	10$\pm$6	9$\pm$5	15$\pm$10	27$\pm$16	36$\pm$17
		10 l	9$\pm$4	10$\pm$6	16$\pm$11	28$\pm$17	38$\pm$16
1001-3000	normal	75 r	9$\pm$3	8$\pm$4	9$\pm$4	12$\pm$6	14$\pm$6
$\bar{x}$ =1985		70 l	8$\pm$4	8$\pm$4	9$\pm$5	12$\pm$6	14$\pm$6
N = 795	slight	25 r	10$\pm$4	11$\pm$7	18$\pm$12	35$\pm$16	37$\pm$15
		30 l	9$\pm$5	10$\pm$6	20$\pm$13	34$\pm$17	38$\pm$16
above 3000	normal	63 r	9$\pm$3	8$\pm$4	10$\pm$5	14$\pm$6	15$\pm$6
$\bar{x}$ = 3993		56 l	8$\pm$4	8$\pm$4	10$\pm$5	13$\pm$7	16$\pm$6
N = 355	slight	36 r	11$\pm$5	12$\pm$7	24$\pm$14	37$\pm$17	41$\pm$19
		43 l	9$\pm$5	11$\pm$8	24$\pm$15	39$\pm$17	41$\pm$18
	marked	1 l	23$\pm$11	39$\pm$16	53$\pm$21	66$\pm$11	66$\pm$13

Table 2

Audiograms of Helicopter Pilots Versus Flying Hours

Flying hours	Category of hearing losses	%		Hearing Losses (dB)				
			1000	2000	3000	4000	6000 Hz
1 - 1000	normal	90 r	9$\pm$3	9$\pm$4	9$\pm$4	11$\pm$5	12$\pm$5
$\bar{x}$ = 438		88 l	8$\pm$4	7$\pm$4	8$\pm$4	9$\pm$5	11$\pm$6
N = 320	slight	10 r	10$\pm$3	10$\pm$4	16$\pm$12	34$\pm$22	41$\pm$18
		12 l	8$\pm$3	9$\pm$5	16$\pm$12	31$\pm$20	36$\pm$19
1001-3000	normal	74 r	9$\pm$3	9$\pm$4	9$\pm$4	12$\pm$5	13$\pm$6
$\bar{x}$ = 1985		69 l	8$\pm$4	8$\pm$5	9$\pm$5	11$\pm$6	13$\pm$6
N = 656	slight	25 r	10$\pm$4	11$\pm$6	21$\pm$15	37$\pm$17	39$\pm$18
		30 l	9$\pm$4	11$\pm$6	22$\pm$14	36$\pm$18	4o$\pm$19
	marked	1 r	11$\pm$8	15$\pm$9	44$\pm$28	67$\pm$19	58$\pm$21
		1 l	15$\pm$12	14$\pm$12	52$\pm$28	59$\pm$29	62$\pm$15
above 3000	normal	58 r	9$\pm$3	9$\pm$4	10$\pm$4	13$\pm$6	15$\pm$7
$\bar{x}$ = 4235		54 l	8$\pm$3	8$\pm$4	10$\pm$5	13$\pm$6	15$\pm$7
N = 395	slight	40 r	10$\pm$5	11$\pm$6	19$\pm$13	35$\pm$17	36$\pm$17
		42 l	9$\pm$4	12$\pm$7	22$\pm$14	38$\pm$17	4o$\pm$16
	marked	2 r	21$\pm$13	33$\pm$18	51$\pm$28	54$\pm$26	53$\pm$25
		3 l	23$\pm$13	37$\pm$17	53$\pm$21	56$\pm$19	47$\pm$18

A comparison of this statistical analysis with a 1970 survey of 2000 pilots revealed a marked decrease in the incidence of hearing losses, especially in senior pilots because: a) World War II pilots finished their flying career, b) inflight noise protection has been improved and c) pilots have become more aware of noise hazards on airfield and shooting range.

A long-term follow-up study has investigated 40 pilots with unilateral impact-noise-induced hearing losses (14 right, 26 left ears) acquired in 1964. In 1980, after 16 more years of age and 3000 more hours of flying time, the affected ears showed no increase in hearing losses at 1000 and 2000 Hz, 14 dB at 3000, 17 dB at 4000 and only 6 dB hearing loss at 6000 Hz. The increase in the non-affected ears was somewhat less: the hearing of all affected ears remained safely within standards. This approves our policy to admit pilot-candidates having a unilateral noise induced high tone hearing loss above 2000 Hz.


Figure 3

CONCLUSIONS

The analysis of pure tone audiograms of 4.034 German military pilots in 1979 revealed, that 73% of total ears have normal hearing up to at least 6000 Hz. 26% have only very slight hearing losses above 2000 Hz with 22 ± 14 dB at 3000 Hz, 36 ± 18 dB at 4000 Hz and 40 ± 18 dB at 6000 Hz. This allows reliable speech communication in every day life and aboard the aircraft. The total of marked hearing losses requiring a waiver was 0.65%, mostly unilateral with prevalence of the left ears and induced by impact noise on the shooting range in the early stages of the career. Thus, from the pilots' side, all higher speech frequencies are available for improved voice communication systems and should be used in high noise environment.

AN ACTIVE NOISE REDUCTION SYSTEM FOR AIRCREW HELMETS

P.D. WHEELER, WOLFSON UNIT FOR NOISE AND VIBRATION CONTROL,
INSTITUTE OF SOUND AND VIBRATION RESEARCH,
UNIVERSITY OF SOUTHAMPTON, SOUTHAMPTON.
SO9 5NH. U.K.   and

S.G. HALLIDAY, WOLFSON INDUSTRIAL UNIT, DEPARTMENT OF
ELECTRONICS, UNIVERSITY OF SOUTHAMPTON,
SOUTHAMPTON. SO9 5NH. U.K.

SUMMARY

An active noise reduction system (ANR) has been developed for use in aircrew helmets, in which the acoustic noise field inside the ear defender is detected using a miniature microphone, and an antiphase signal is fed back to a communications telephone within the ear defender. The objectives for this development have been to improve speech intelligibility, and to reduce noise exposure.

Developed under contract to the UK Ministry of Defence (Procurement Executive), by the University of Southampton, the ANR system has been comprehensively tested in a series of laboratory trials, followed by flight trials in a number of aircraft.

In the laboratory trials, a group of eighteen subjects, wearing an ANR-modified helmet, were exposed to an external noise field similar to that experienced by aircrew in a high performance strike aircraft. Comparisons of attenuation and speech intelligibility scores, with and without the ANR system in operation, were made. The modified helmet's performance was also compared to that of the standard helmet.

In order to eliminate user controls, an adaptive control facility has recently been added, which optimises the degree of noise cancellation to compensate for variations in helmet fit between wearers.

This paper describes the development of the active noise reduction system, and presents the results of laboratory and flight trials.

INTRODUCTION

The cockpit noise environment of current generation strike aircraft is typically of the order of 110-120 dB overall sound pressure level, with highest levels in the 250 and 500 Hz octave bands. Although in some aircraft types, cockpit air conditioning systems and/or discrete frequency engine sources produce substantial levels at higher frequencies (typically in the 1, 2 kHz bands),and overall levels are themselves dependent upon flight profile, in general terms, a typical in-flight cockpit noise spectrum might take the form shown in Figure 1.

Now the attenuation provided by an aircrew helmet is generally poor at low frequencies, in the 250 Hz region, but rises rapidly above 500 Hz, as indicated in Figure 2 (Reference 1).

Thus combining these data, a typical noise spectrum at the crew member's ear will be entirely low-frequency dominated, having an extremely steep spectrum slope of the form shown in Figure 3.

The generalised frequency response (in muff) of the communications telephone of the helmet is a rising characteristic from 150-1500 Hz (at 20 dB/decade), and a rapidly falling characteristic (40 dB/decade) above 1500 Hz. Combining this with the idealised speech spectrum of the male voice, the typical speech spectrum delivered to the crew member's ear is that of Figure 4. Superimposing Figure 4 on Figure 3 shows that if the overall speech level is set for a 10 dB signal-to-noise ratio at 1-2 kHz, there is negative S/N at 500 Hz (without including the effects of upward-spread of masking).

Therefore, if an improvement in low-frequency attenuation could be achieved for the helmet, not only would there be a direct reduction in total noise level at the ear, but also a direct reduction in overall A-weighted level, with its attendant reduction in hearing-damage risk, and a potential improvement in speech intelligibility. This then was the starting point for the development of an active noise reduction system, with a comparatively modest initial target of 10 dB additional attenuation over the range 300-800 Hz as specified by MOD(PE) UK.

PRINCIPLES OF OPERATION

A block diagram of the system is shown in Figure 5. It comprises a negative feedback loop, incorporating an acoustic path between the telephone and microphone. A complete analysis of the system shows that the sound pressure at the ear (P_e) is comprised of two components due to the signal (S_v) and the external noise field. If the noise field within the muff without the ANR loop closed is N_p , then with ANR in operation

$$P_e = \frac{(AT_1)}{1 + ABKT_1T_2} \cdot S_v + \frac{(1)}{1 + ABKT_1T_2} \cdot N_p \qquad \ldots \ldots (1)$$

where A, B, K, T_1, T_2 are the transfer functions of the amplifier, feedback loop, acoustic cavity, telephone, and microphone respectively.

From (1) it can be seen that the amplification of the two components is different. If the signal S_v is pre-emphasised and returned to its original level then an improvement of a factor $(1 + ABKT_1T_2)$ can be achieved for the noise at the ear. The product KT_1T_2 is of fundamental significance in that it represents the overall transfer function of the telephone-cavity-microphone combination. This electroacoustical response dominates the overall response of the system, and unfortunately varies from wearer to wearer (and to a lesser amount from wearing to wearing) since it is in part associated with the fit of the muff over the wearer's ear. To provide full cancellation at all frequencies, the product AB would have to be exactly complementary to KT_1T_2 in terms of amplitude and phase, at all frequencies. In practice, this degree of equalisation cannot readily be achieved, without sacrificing speed of response, even if KT_1T_2 could be held constant for all wearers, and so cancellation is only obtained over a limited range of frequencies, typically 50 Hz to 2 kHz.

The design of the ANR-modified ear defender is based upon the existing ear defe- - shell and seal of the UK Mk4 aircrew helmet. The transfer function of the existing telephone of the helmet was not suitable for application to ANR, and a moving-coil dev :e has been used instead. The microphone used in the feedback loop is a miniature electret type.

From equation (1) it can be seen that if $|1 + ABKT_1T_2| < 1$ then noise enhancement is predicted. This is observed at high frequencies (with current development typically 5 kHz) where the product AB is no longer complementary to KT_1T_2. Such frequencies are outside the main information-carrying band of speech, and since the helmet provides some 30-40 dB passive attenuation, some small increase in noise level was regarded as acceptable. Such components of speech that are transmitted by the communications system at this frequency are also boosted, by an equal amount, by the process of enhancement.

As has previously been shown, the amount of noise reduction achieved is dependent upon the gain AB of the feedback loop. As the maximum gain which can be applied before instability occurs depends upon the accuracy with which AB matches KT_1T_2, the system gain has to be set for each wearer, in order to obtain maximum potential from the system. Since KT_1T_2 also varies between left and right ears for most wearers, two entirely independent ANR feedback loops are used, and therefore the setting-up procedure requires the setting of a gain control for each ear.

In order to eliminate this setting-up requirement, an adaptive gain control was developed for the ANR system, before extensive trials in single-seater strike aircraft were undertaken. This circuit monitors the noise level at the ear, and continually adjusts the feedback loop gain to maintain a minimum noise condition in-flight. In this way, not only are inter-wearer gain changes eliminated, but the small variations in helmet fit (and therefore in KT_1T_2), which occur with high -"g" in-flight manoeuvres, are also compensated for.

The original design specification, referred to earlier, was for an additional attenu- ation of 10 dB over the range 300-800 Hz. The results of laboaratory and flight trials show that this has been met and exceeded, with the current kit producing a typical performance of 15-20 dB active attenuation, and maximum attenuations of up to 25 dB, in a working range extending from below 50 Hz to beyond 2.5 kHz.

LABORATORY TRIALS

In 1977 a definitive series of laboratory trials was commenced in order to quantify the ANR system's performance for a large number of subjects prior to starting flight testing. During these trials, the performance of the ANR system was measured in a simulated aircraft noise environment. Eighteen subjects were exposed, in turn, in a reverberant chamber, to the external noise field shown in Figure 6, which is typical of a Buccaneer strike aircraft, in low level high-speed flight, while objective and subjective tests were carried out.

The objective test comprised the measurement of the sound pressure level at the subjects ear (i.e. in the ear defender), using a Knowles miniature microphone. From these data, the degree of active noise reduction was determined from the difference in ANR on and off levels, and the passive attenuation of the ear defender from the external and at-ear levels. As a comparison, the attenuation of the existing Mk4 helmet was measured using the same technique.

The subjective test comprised an intelligibility test using Anglicised Modified Rhyme Test word sets. Subjects scored word reception in noise for all three conditions, i.e. existing Mk4 helmet (and telephone), and ANR-modified helmet with ANR in and out of operation, and to eliminate differences due to variations in frequency response between the two telephones, the speech spectrum delivered to the ear was balanced to ±1 dB overall, and ±1 dB in each 1/3 octave band from 100 Hz to 5 kHz, by spectrum shaping, for all

conditions. The word score test then became a subjective measure of the change in noise field at the ear brought upon by ANR. A balanced experimental design was employed, in order to minimise inherent errors in the source word set matching.

Finally, the source speech material was degraded by mask-microphone "pick-up" noise appropriate to the experimental noise conditions, and the word score test was repeated.

Results for passive attenuation showed that the ANR-modified ear defender was not inferior to the Mk4 defender in performance. Mean values (both ears) are shown in Figure 7.

Mean ANR performance (both ears) is shown in Figure 8, for the eighteen subjects, indicating a plateau of 14-15 dB active attenuation, with a working range extending from below 100 Hz to above 1.6 kHz. At frequencies above 2 kHz some enhancement of noise was observed, averaging 3 dB up to 5 kHz.

The word score tests showed a mean improvement in score of 26%, starting from a base 50% score, for ANR in versus out of operation, for "clean" speech material, and an improvement of 21%, from 50%, for speech "degraded" by mask microphone effects.

These increases in word scores were approximately consistent with the predicted increases based upon the Articulation Index procedure.

INITIAL FLIGHT TRIALS

Following the laboratory trials described above, three flyable ANR kits were built, in a form suitable for installation in a Buccaneer aircraft, the intention being to carry out a series of flight trials using MOD(PE) staff as trial subjects. The kits were designed to integrate with the aircraft power supply, and provided a number of data outputs for in-flight monitoring of system behaviour via a multi-track tape recorder. Unfortunately, before these trials could be commenced, the aircraft concerned was withdrawn from trial work.

The opportunity was then taken to test the ANR system in a transport aircraft (a Hercules), enabling a number of subjects to take part simultaneously, and allowing University staff to experience the operation of the kit at first hand, in flight. During such flights, subjects made individual subjective assessments of the performance of the kits in terms of noise reduction, and effect upon communications, and in-muff noise recordings were made, using a Nagra IV SJ portable recorder. Subjects disconnected from the aircraft CCS for recordings, in order to obtain interference-free data. Figure 9 shows the mean ANR performance obtained in these flight trials, with plateau attenuation values of 16-17 dB, over a working range from below 50 Hz to 2.5 kHz.

MINIATURE ANR FLIGHT KITS

With the success of the early flight trials, further trials were planned in strike aircraft, with RAF personnel as trial subjects. A prototype miniaturised kit was designed and built, some 150 x 100 x 25mm in size. The kit was battery powered, and intended to be worn in the knee pocket of a flying suit. Data outlets were provided for in-flight recordings on a miniature Nagra SN tape recorder. Nine sorties were flown with this kit by an RAF navigator at the Aeroplane & Armament Experimental Establishment, Boscombe Down, UK, in Hawk, Harrier, Jaguar, Phantom, and Tornado aircraft.

Flight conditions varied from high level handling, through low level navigation, to simulated combat. In addition to subjective observation and assessment by the navigator, a large number of in-flight recordings were made, and later analysed. The mean ANR performance, over all aircraft types, for the trials was a plateau of 14-15 dB active attenuation with a working range of 50 Hz - 2.5 kHz. Subjective assessment noted a substantial improvement in noise level at the ear.

Following this, a quantity of miniature kits were built for more extensive user-evaluation by RAF aircrew. The kits were again battery-powered and small enough to fit in a flying suit pocket. Most of the kits incorporated an adaptive gain control facility, as described earlier.

At the time of writing, these trials are in progress, and final results cannot be quoted. However, initial response from the aircrew corroborates the subjective assessments of earlier subjects, and the recorded data show ANR performances similar to those quoted above, attenuations of up to 20 dB being measured. Figure 10 illustrates a typical performance.

CONCLUSIONS

The viability of the ANR system has been proved during laboratory and flight trials in a range of aircraft. Continual development of the electroacoustical design of the system has allowed a steady improvement in performance to be made, such that the current design provides some 15-20 dB active attenuation over a working range from below 50 Hz to beyond 2 kHz. The effect of this additional attenuation is to reduce in-flight noise levels at aircrew members ears by some 15-20 dB(A), depending upon aircraft type.

The system could be incorporated into aircraft communications systems at the design stage, or could be used in man-pack form in the short term. It is intended to investigate

the application of the system to rotary-wing aircraft in the near future.

REFERENCE

1.   G.R. ROOD,   Unpublished MOD(PE) Report.

ACKNOWLEDGEMENTS
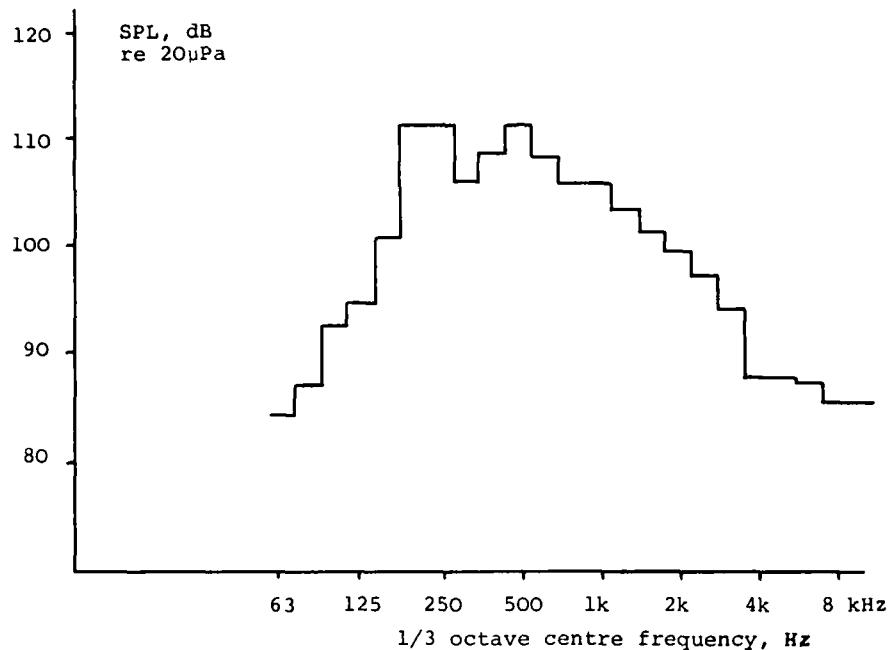
FIGURE 1: TYPICAL IN-FLIGHT COCKPIT NOISE SPECTRUM,
STRIKE AIRCRAFT
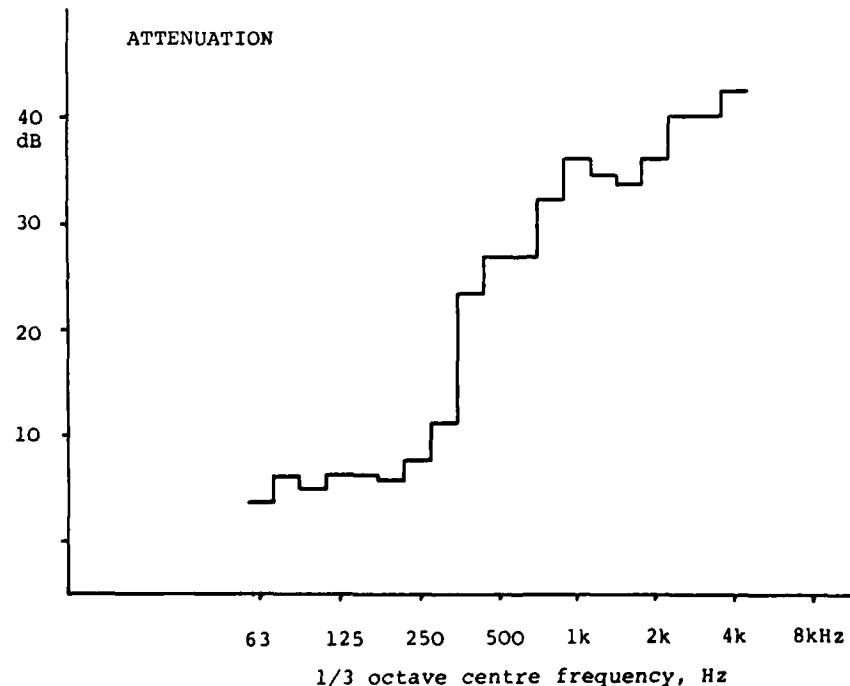
FIGURE 2: Mk4 HELMET - INSERTION LOSS (MEAN OF 60
MEASUREMENTS)

SPL dB
re 20μPa

1/3 octave centre frequency, Hz
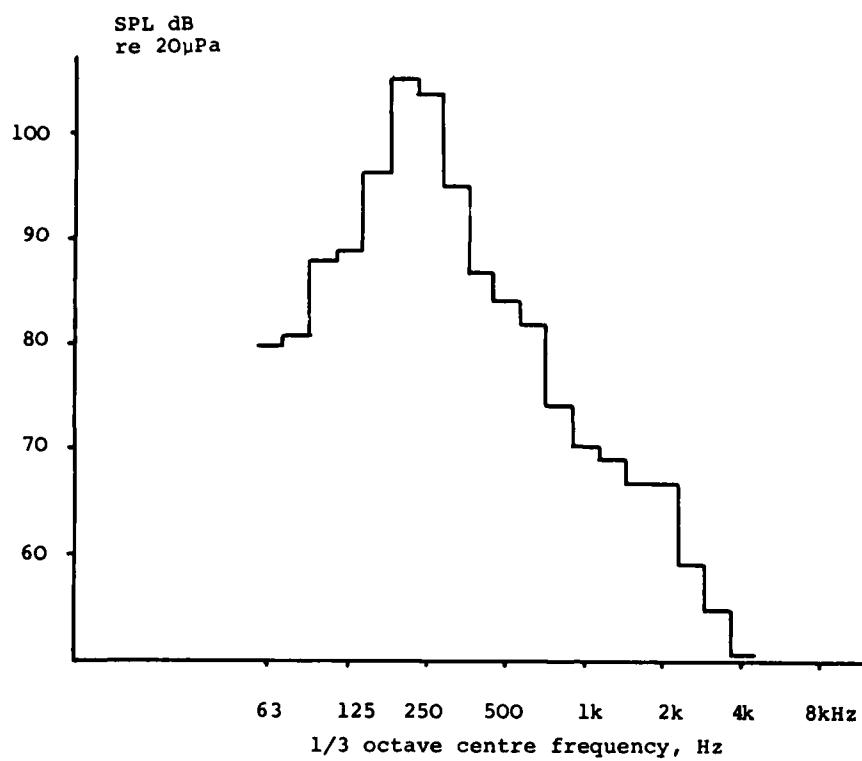
FIGURE 3: SPL AT CREW MEMBER'S EAR

SPL,
ARBITRARY ZERO

SPEECH ENVELOPE
LIMITS

1/3 octave centre frequency, Hz
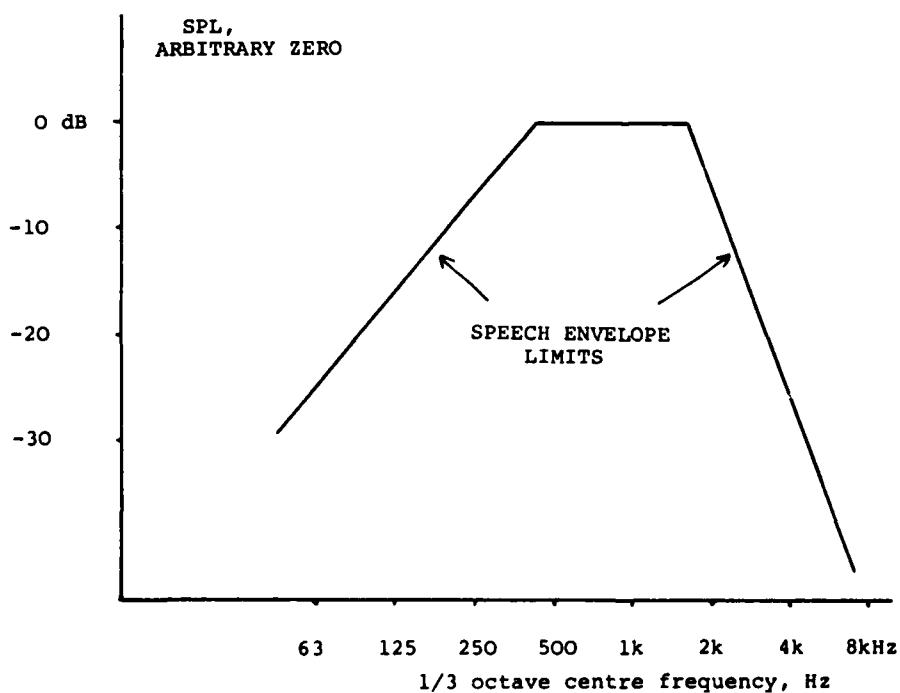
FIGURE 4: IDEALISED SPEECH SPECTRUM ENVELOPE AT EAR

NOISE

SPEECH
IN

SPEECH
PREAMP

A

TEL

MIC

EAR DEFENDER

B

FEEDBACK FILTER
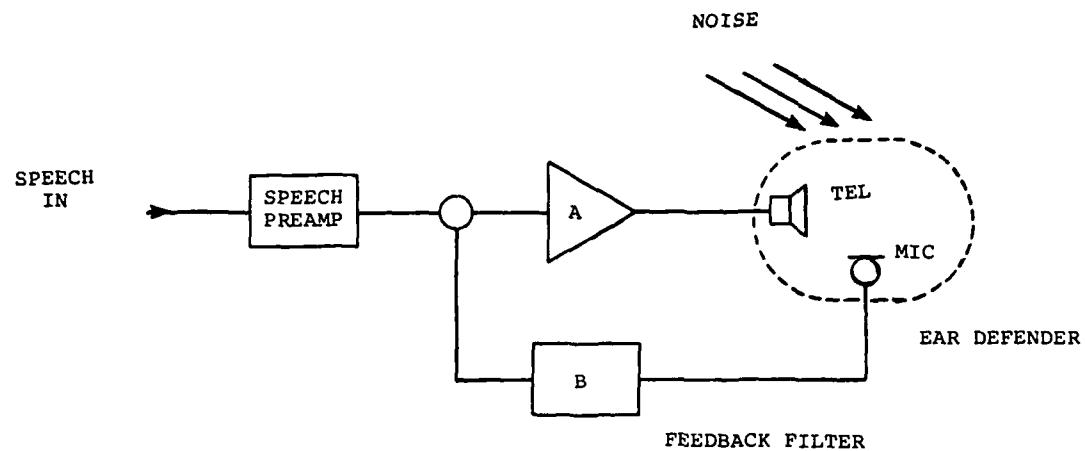
FIGURE 5: SYSTEM BLOCK DIAGRAM

SPL, dB
re 20μPa

110

100

90

80

70

63   125   250   500   1kHz   2k   4k   8kHz

1/3 octave centre frequency, Hz
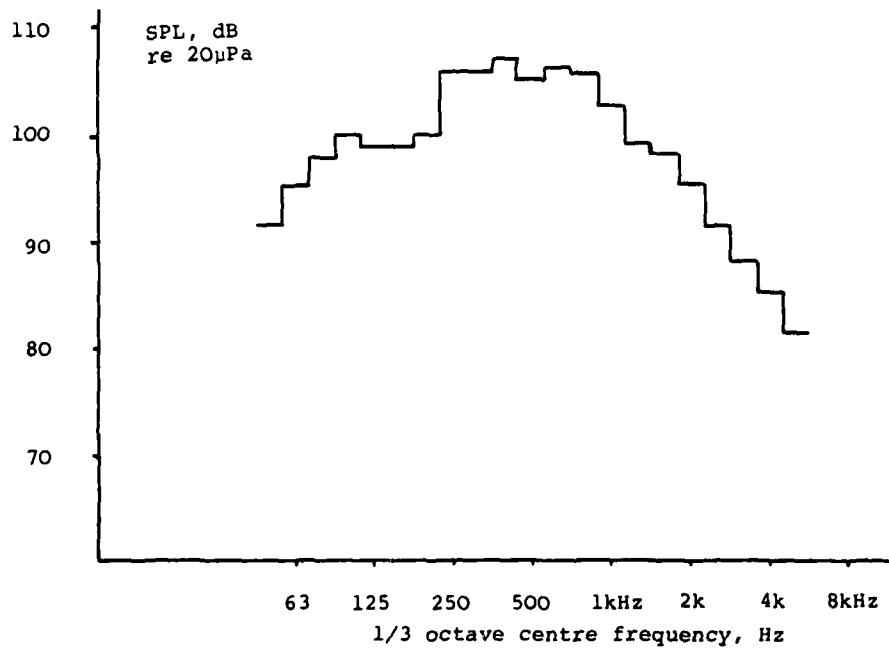
FIGURE 6: LABORATORY TEST NOISE SPECTRUM

ATTENUATION

ANR - MODIFIED HELMET

STANDARD HELMET

63   125   250   500   1k   2k   4k   8 kHz

1/3 octave centre frequency, Hz
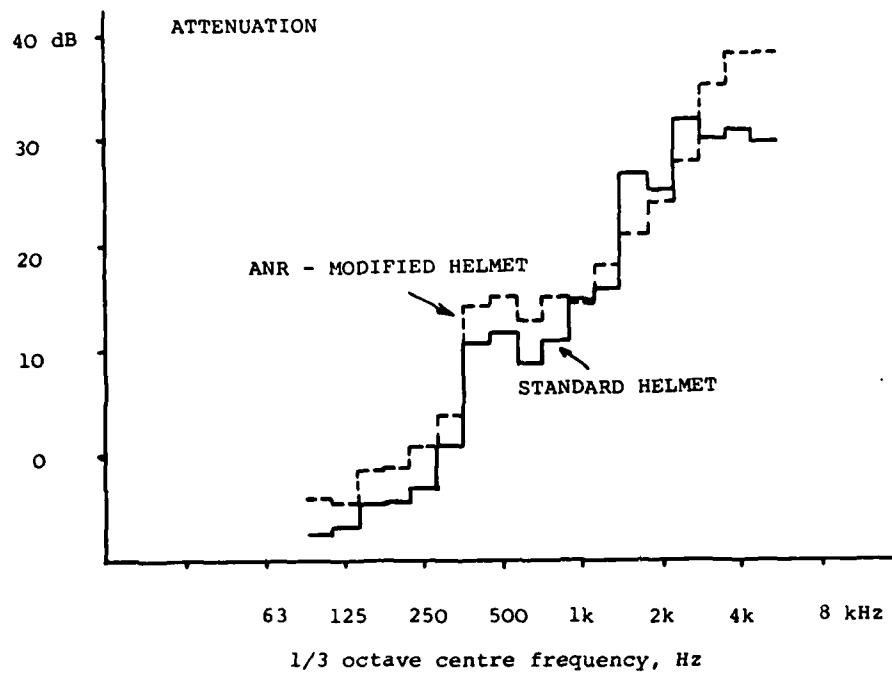
FIGURE 7: ATTENUATION OF STANDARD vs MODIFIED HELMETS

ACTIVE ATTENUATION

63   125   250   500   1k   2k   4k   8kHz

1/3 octave centre frequency, Hz
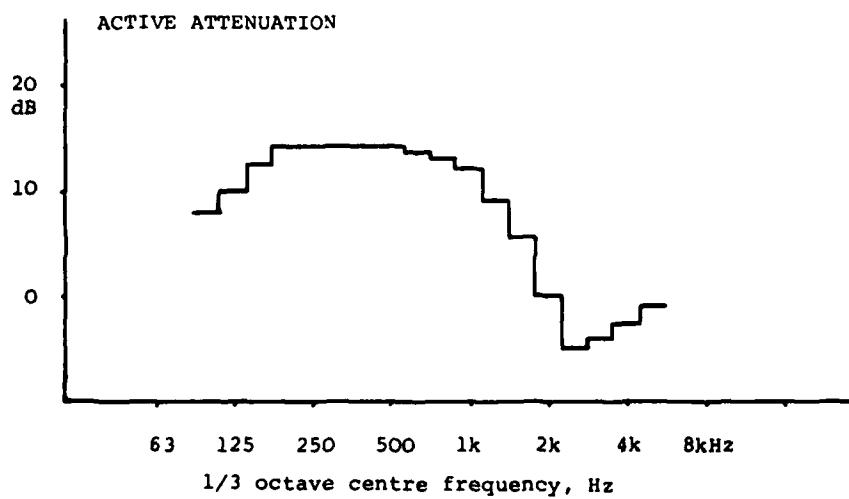
FIGURE 8: MEAN ACTIVE ATTENUATION IN LABORATORY TEST

ACTIVE ATTENUATION
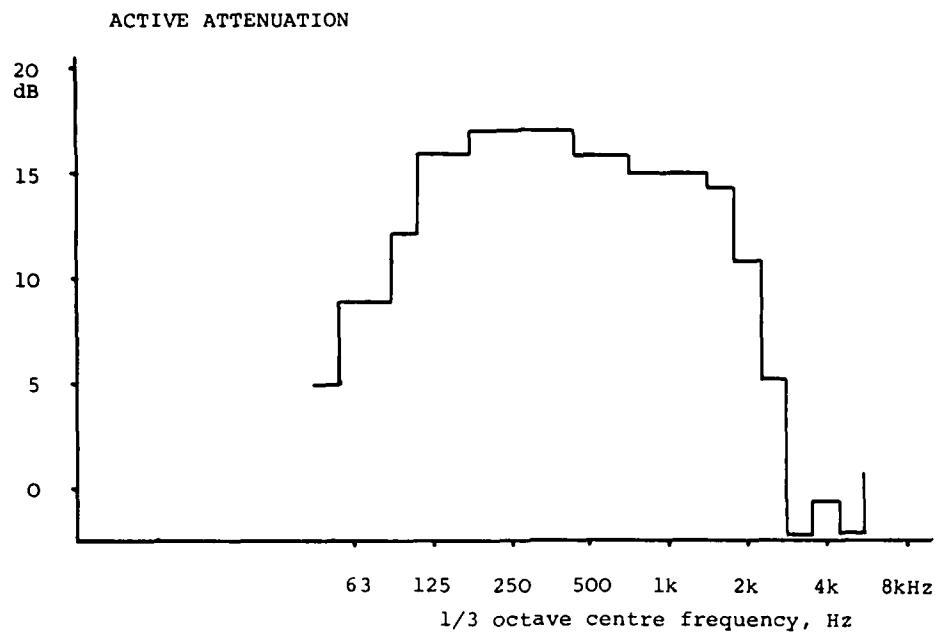
FIGURE 9: MEAN ACTIVE ATTENUATION IN HERCULES TRIALS
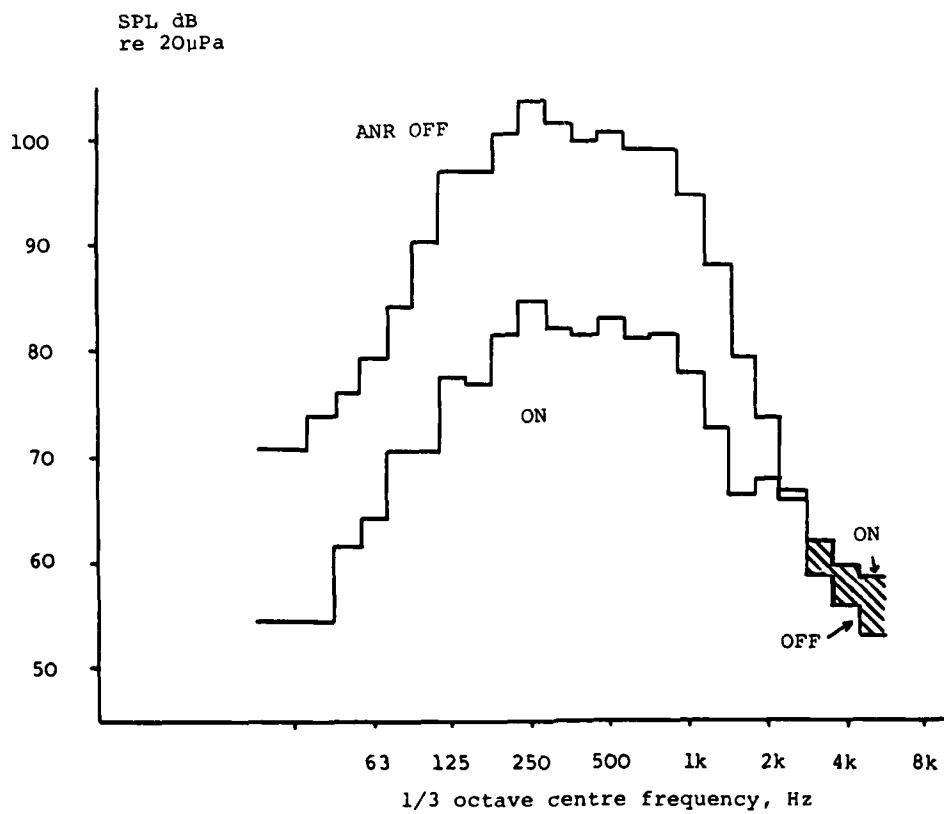
FIGURE 10: SOUND PRESSURE LEVEL AT CREWMAN'S EAR
BUCCANEER 420 KNOTS LOW LEVEL

HEARING IMPAIRED AVIATORS IN THE US ARMY

BY

Susan Liff, CPT, MSC
US Army Aeromedical Center
Fort Rucker, AL   36362
USA
Jerod L. Goldstein, LTC, MSC
US Army Aeromedical Research Laboratory
Fort Rucker, AL   36362
USA

SUMMARY

This article presents an audiometric profile of a group of US Army aviators who had failed to meet the minimal acceptable hearing loss standard and were granted permission to continue to fly under medical waiver. The flight safety records of this group were evaluated to determine if any relationship existed between hearing loss and flight mishaps or accidents involving these individuals.

Introduction and Background

The US Army has established standards of medical fitness for Army Aviation. Included in these standards is an acceptable minimal hearing level which must be maintained to remain on flight status. The reasons for a hearing loss standard are two-fold: one to establish criteria beyond which an aviator would not be considered safe to fly, and two, to provide a preventive medicine function by grounding an aviator prior to his suffering significant irreversible hearing loss.

In evaluating an aviator who exceeds hearing loss standards, consideration is given to the following:

1. At what point would communication become difficult and cause a safety hazard?

2. What are appropriate diagnostic test procedures to evaluate the aviators hearing ability?

3. At what point does flight experience cease to compensate for the hearing loss?

In evaluating the aviator for waiver or removal from flight status, a composite picture of the aviator must be considered to include safety, preventive medicine, flight experience, duration the condition has existed, stability of the condition and the needs of the Army.

The US Army has established a retention standard for hearing loss as its Class II levels identified in AR 40-501 change 31 which are as follows:

	5000Hz	1000Hz	2000Hz	4000Hz
Better Ear	35x	30	30	-
Poorer Ear	35	50	50	-

x Hearing levels in decibels (dB) International Standard Organization (ISO) 1964 Standard.

If the aviator's hearing loss levels exceed this standard, his case is reviewed on an individual basis by the Aeromedical Consultant Advisory Panel (ACAP) which is composed of a group of Aerospace Medicine Specialists (flight surgeons). In reviewing each case, the panel requests an audiological evaluation consisting of at least pure tone thresholds, speech reception threshold testing, speech discrimination testing in quiet, and a measure of middle ear function which could consist of bone conduction testing but usually involved impedance audiometry. The panel also requests past audiograms to evaluate the progressive nature of the hearing loss and a statement from the individual and his local flight surgeon which includes a record of flight experience. The ACAP only reviews the medical records of aviators whose cases are brought before it. That is, if a local flight surgeon feels there is just cause to remove the aviator from flight status, the ACAP will usually concur and not review the case in any depth.

With the advent of certified audiologists in the US Army and programs to improve screening testing procedures such as used in the annual flight physical, an increased number of aviators may be found to have hearing losses which exceed the Class II level. To date, of the 12,680 aviators and some 500 civilian instructor pilots and aviators, 70 (0.5%) are flying under a medical waiver for hearing loss. This figure does not represent the suspected number of aviators (active duty and civilian), but is the known population currently on waiver.

This paper will describe the audiologic profile of this population of 70 aviators to include their mishap statistics as cataloged by the US Army Safety Center. It will attempt to determine if there is any relationship between their hearing loss and accident rate and if the method of evaluation used by the ACAP is an appropriate and sensitive method.

The waiver records of the ACAP were reviewed and a total of 70 aviators were found to be on waiver and still currently flying. Also, all audiometric data was available on these individuals to provide the necessary profile of this group.

The mishap and safety records of these individuals were reviewed at the US Army Safety Center to determine if any individuals were involved in a mishap or accident while flying and if hearing loss (communication) was found to be a causal factor.

Findings

Table I presents some descriptive statistics for the group to include their age at the time the study was conducted, the number of years they have been employed by the government, the years of government service they had served at the time their waiver was granted, how many years they had been flying under the waiver at the time of the study, the total number of hours they had flown at the time the study was conducted, and the total number of hours they had flown at the time their waiver was granted. Table II presents the pure tone audiometric test findings and speech discrimination test findings. Of the 70 aviators, the US Army Safety Center had records of three mishaps where the cause involved pilot error. In all three cases the mishap occurred prior to the time the waiver was granted to the aviator. In the one case where communication was a causal factor, the aviator was involved in the mishap in 1977; while his waiver was granted in 1979. The aviator had flown a total of 12,646 hours up to 1979 when the ACAP reviewed his case and he accumulated an additional 400 hours from 1979 through 1980.

Accident investigation statistics were obtained from DA Form 2397 line 9 which identifies communication problems as a causal factor such as misinterpretation, language barrier, noise interference and others. In the first two cases, none of these factors were indicated but in the third case line 9 was checked as a contributing factor. Specifically the report indicated the Instructor Pilot (IP) was getting ready to show a new maneuver to his student. Prior to beginning the maneuver, the IP tried to contact the tower. There was no response from the tower---therein lies the communication problem. The IP, rather than waiting for a response, went ahead and showed the maneuver, which resulted in failure. The Board went on to say that it was "overconfidence" of the IP that contributed to the mishap.

Discussion

Seventy hearing impaired aviators under medical waivers in excess of Class II hearing standards are currently on flight status. A composite mean picture of this group shows an aviator with 21 years of government service, 44 years of age, 5252 flight hours and being on a waiver for three years. Although, in general, we can say that these aviators are a very experienced group, some of the population are as young as 25 years of age, with only three years of government service and only 240 flight hours. Since being granted a waiver, the group has flown an average of 675 hours with as little as no hours and as many as 2000 hours. The mean pure tone thresholds indicate the typical noise induced high frequency sensori-neural hearing loss with variation above or below this mean level. Although the mean speech discrimination scores for the group was 89% for the right ear and 86% for the left ear, some of the speech discrimination scores were as low as 52%. According to the mishap statistics maintained at the US Army Safety Center, only three of these 70 aviators were involved in mishaps attributed to pilot error. Only one of these appears to have been attributed to a communications problem in that it was due to a nonstandard maneuver being performed prior to getting a clearance from the control tower. It would be difficult to prove that the hearing loss of the pilot presented the communication problem. That is, there was no communication for the pilot to hear. This particular aviator although below Class II standards has hearing better than the man of the group at 2KHz and worse than the group at 3-8KHz. His audiogram in dB ISO hearing levels (HL) was as follows:

	0.5KHZ	1KHZ	2KHZ	3KHZ	4KHZ	6KHZ	8KHZ
Right	10	15	36	65	90	NR	NR
Left	5	15	35	70	100	NR	NR

Speech discrimination was 96% right, 80% left.

This group of hearing impaired aviators appears to be performing their aviation assignments safely, as measured by the mishap statistics of the US Army Safety Center. Although experience in aviation can certainly be expected to help overcome auditory difficulties, some of these aviators have under 1000 hours of flight time but this limited experience does not appear to be detrimental.

It appears that in the case of these 70 aviators the ACAP has chosen correctly in allowing them to have a waiver to fly despite their hearing loss. A point at which suspension from flying should occur was not noted from the audiologic statistics of this group. While you certainly do not want to waive any aviator who would jeopardize his own or another's safety with his inability to hear, you also do not want to suspend an aviator who is safe to fly. Some of the aviators in this study rank among the more experienced in the Army. We know that we have at least 70 hearing impaired aviators flying with a waiver who are safe flyers.

Conclusion

1. No relationship seems to exist between the pilot error (accident rate) of the aviator and a hearing loss level above Class II standards for 70 Army aviators.

2. The method of evaluation used by the ACAP for these 70 aviators appears to be sensitive enough to recognize who is safe to continue flying.

3. The pure tone test results and the Class II hearing levels seem to be a good starting point at which to perform a detailed audiologic evaluation of the hearing loss aviator. But due to the wide range of speech discrimination scores and the poor performance by some in this type of testing, it does not seem to be sensitive enough in recognizing communication problems the aviator may experience. Therefore, other types of inflight or aviation specific speech communication tests need to be developed and standardized with the Army aviation population to provde this needed specificity.

REFERENCE
Department of the Army. 1960. Medical Service Standards of Medical Fitness. Washington, DC: Department of the Army. AR 40-501.

TABLE I

Descriptive Information of 70 Aviators*
Waivered for Hearing Loss as of 1980

	Age in Yrs at time of study	Yrs of Gov service at time of study	Yrs in Gov service at time of waiver	Yrs since waiver was granted	Total hrs flown at time of study	Total hrs flown at time of waiver
Mean	44	21	19	3	5252	4577
Standard Deviation	9	8	8	2	4144	3827
Median	42	20	17	2	3025	3000
Mode	41	23	13	1	3500	3000
Range	26-63	3-37	2-35	1-11	240-18800	75-17880

* Includes US Army and Department of the Army Civilians flying Army aircraft primarily helicopter.

TABLE II

Hearing Loss in Decibels (dB)* by Frequency and
Speech Discrimination Test Results of Seventy Aviators
Waivered for Hearing Loss as of 1980

RIGHT EAR	500	1K	2K	3K	4K	6K	8K	Speech Discrimination Testing
Mean	13	16	40	54	55	51	43	89
Standard Deviation	14	14	17	20	23	26	23	11
Median	10	10	40	50	55	45	25	92
Mode	5	10	55	55	75	20	65	96
Range	0-70	0-55	0-75	10-100	5-100	0-100	0-95	52-100
LEFT EAR								
Mean	13	19	51	60	61	58	48	86
Standard Deviation	14	15	15	19	21	21	23	12
Median	10	15	55	60	60	60	30	88
Mode	10	15	65	55	56	60	85	92
Range	0-60	0-60	10-70	10-105	10-110	10-100	0-80	52-100

* Levels expressed reference ISO standard 1964.

HEARING CONSERVATION

Dr. S. Kanagasabay M.Sc, MBBS, MFCM, MFOM,
Senior Medical Officer, Research & Development (1),
Directorate of Civilian Medical Services (PE),
Ministry of Defence,
Empress State Building,
London SW6 1TR,

SUMMARY

Impairment of hearing is one of the adverse effects of noise. The measures adopted to minimise hearing impairment, conveniently termed Hearing Conservation, include specifying an acceptable noise exposure, identification of personnel at risk, provision of suitable protective equipment, limitation of exposure time where this is necessary and medical monitoring by audiometry. This paper is based on proposals in the United Kingdom, for noise criteria and for audiometry in industry.

INTRODUCTION

Noise has been defined as sound which may cause undesirable effects. These effects include;

. interference with communication
. disturbance of work, recreation or sleep
. impairment of hearing
. general body effects at the highest intensities.

This paper deals only with the prevention of hearing impairment. The preventive measures rest on the principle that noise must be measured in order to make an assessment of its possible effects on hearing. Previous research has indicated the relations between noise and hearing, so that hazardous situations can be identified on the basis of noise measurements. The effect of the noise is primarily determined by the total noise energy received by the ear, this quantity is denoted by the term noise exposure, and is the product of the intensity of the noise and its duration. Since the hazard from a particular exposure is known, and the intensity and probable duration of the noise over a working lifetime can be estimated, it follows that an acceptable maximum of noise can be specified. This has been variously described as; limiting sound level, exposure standard, recommended standard and so on. Where this level is exceeded exposure may be restricted by reduction of noise at source. If this is not possible or practical, other measures must be used to reduce the exposure of employed persons to noise. This may require the use of ear protection and, at the higher levels limitation of duration of exposure may also be necessary.

In the United Kingdom, the question of noise was considered at a national level in 1960 with the appointment of the Committee on the Problem of Noise. The report of this committee was submitted to parliament in 1963, (1). In compiling this report the committee took advice from specialists, members from scientific and industrial organisations, local authorities and associations of citizens of various kinds. It also took account of the relevant published work on; General effects of noise on people and, Occupational exposure to high levels of noise. I would like to dwell briefly on that part of the report dealing with occupational exposure to high levels of noise, as it is relevant to hearing conservation. This was considered under the following headings;

. effects on the ear
. criteria for safeguarding hearing
. ascertainment of individual susceptibility
. reduction of noise exposure.

The effects on the ear were related to temporary threshold shift and permanent threshold shift though, at that time, there was no attempt to relate the two. It was assumed that presbycusis and hearing loss due to noise or permanent threshold shift were additive. Permanent threshold shift was, in most cases, seen to occur as a small loss at 4 kilohertz (kHz) and that adjacent frequencies are affected as the condition progressed. The relationship between noise exposure and threshold shift was governed by many variables which include sound pressure level, distribution of sound energy over the frequency spectrum, the duration of exposure and its distribution in time, individual susceptibility and changes in hearing threshold due to increasing age.

In looking for criteria for safeguarding hearing a number of investigations conducted into the relationship between noise and hearing loss were considered. On the basis of this noise levels were suggested at which measures to protect the individual against hearing loss should be undertaken. These were similar to the levels suggested by experts in the U.K. and the U.S,A. for long exposure to broad band noise. It was also suggested that the total energy incident on the ear over a given time determines the possible occurence of hearing loss. On this basis each doubling of the energy (that is the addition of 3 dB) could be compensated by halving the duration of exposure. This principle which is know as the equal

energy assumption was not thought, at that time, to be applicable to durations of exposure less than about one hour. The equal energy assumption is still the subject of much discussion.

Noise levels and data on the extent of hearing loss in given circumstances are generalisations and do not allow the accurate prediction of an individual exposure to noise. This is because of the variation in susceptibility between persons. This being so it was felt that some means of obtaining an early indication of the susceptibility of individuals to hearing loss in any given noise environment would be useful. Having considered the available evidence the view taken was that individual susceptibility to permanent threshold shift can only be found by periodical audiometric examination.

For purposes of reduction of the noise hazard it was agreed that there were two courses of action;

- to reduce the noise to levels which are safe or reasonably so,
- or to protect the ears.

In its conclusions and recommendations the committee recognised that this hazard from noise had been appreciated in some countries and, in the U.K., in some industries and in the Services. In the Royal Navy and in the Royal Air Force there were hearing conservation programmes based on regular audiometric examinations of persons who are exposed to noise which may be hazardous, and upon minimizing of exposure as far as possible, by reducing the noise at source and by protecting the ear. However, the noise in many other working environments remains hazardous and little was being done to investigate the degree of hazard and minimize it. Because of this the following lines of attack on the problem were suggested;

- more widespread voluntary action within industry on the basis of existing knowledge
- legislation on the basis of existing knowledge
- research to try to obtain a more definite understanding of the relationship between noise and hearing loss, with legislation to follow, if necessary, when the results of the research are available.

Additionally, several other recommendations which would serve as interim measures were made; as stated previously these were made in 1963. I have outlined this section of the report of the Committee on the Problem of Noise - also referred to as the Wilson Report - because I believe this was a significant step in dealing with noise as an occupational health hazard at a national level, and the recognition for the need to institute hearing conservation programmes. With the foregoing in mind it would be appropriate to look at the international scene before considering the current proposals in the U.K.

INTERNATIONAL APPROACH

The American Speech and Hearing Association organised a conference on Noise as a Public Health Hazard in 1968. Its aim was to bring together a group of speakers who could present summaries of the current state of knowledge on all aspects of the problem of noise. It was evident that noise pollution was going to be a major issue particularly as the environmentalist movement gathered momentum. Because of this and the other problems which needed further study a permanent committee of ASHA was established, one of whose charges was to plan another congress. Such a congress was held in Dubrovnik in 1973, and the third congress was held in Freiburg in 1978. The five year cycle was deemed an appropriate interval. This was based on the formation in 1973 (2), of five international teams -later increased to eight - who were to accumulate and coordinate knowledge about the effects of noise under the following headings;

- noise induced hearing loss
- noise and communication
- non-auditory physiological effects induced by noise
- the influence of noise on performance and behaviour
- noise disturbed sleep
- community response to noise
- wild life and noise
- effects of interactions between noise and other physical and/or chemical agents.

The team pertinent to the subject of my paper is that dealing with noise induced hearing loss. In 1973 Ward (2) in his preview of the congress content attempted to clarify the position with regard to the term criterion especially when used with qualifiers, such as, noise criteria, or noise exposure criteria. He also outlined the host of questions which needed to be answered. These included the following;

- is hearing above 3 kHz important to the perception of speech? If so under what conditions?
- what frequency weighting scheme, such as A-weighting and D-weighting, gives the closest prediction of the speech-masking ability of a noise?
- can there be damage to hearing without a change in sensitivity?
- what single exposure (8 hour or less) will just produce a significant permanent threshold shift (PTS)?
- what relatively steady state noise exposure, 8 hr/day, for many years, will just produce PTS that exceeds that ascribable to presbyacusis plus sociacusis?
- is there any way to correct audiometric data for presbyacusis plus sociacusis other than simple (and probably incorrect) subtraction?
- under what conditions does the equal-energy hypothesis hold for steady state exposures?
- can individual differences in susceptibility to PTS be predicted?
- can this susceptibility be changed by drugs or diet?

These and many other questions were unanswered at the congress. Five years later, in 1978, Ward (3) as co-chairman of the team reviewed noise induced hearing loss since 1973. He said that it would take him more nearly 20 hours than the 20 minutes allocated to review adequately all the work done in the five year period. However, one fact emerged and that was that there had been no major break through in

in the field of noise induced hearing loss. Most of the major issues which existed five years ago and even before still remain in dispute. Human experiments on temporary threshold shift are increasingly viewed as irrelevant by some researchers. Additionally, various national committees on the ethical oversight of human experiments do not favour this approach because industrial noise exposures are not supposed to exceed 90 dB(A) for eight hours. He felt that the best approach was in the performance of audiometry on a large scale in most industries and indeed in some entire countries, provided these are conducted carefully and accompanied by some completely standardised questionaire; a view I entirely support. You will recall that this does not vary with the proposal made by the Committee on the Problem of noise in 1963.

CURRENT PROPOSALS

In 1972 the Department of Employment of the United Kingdom Government published a ode of ractice for Reducing the Exposure of Employed Persons to Noise'(4). This laid down limiting sound levels based on A-weighting as follows;

Continuous exposure

If exposure is continued for 8 hours in any one day, and is to a reasonably steady sound, the sound level should not exceed 90 dB(A).

Non-continuous exposure

If the exposure is for a period other than 8 hours, or if the sound level is fluctuating, an equivalent continuous sound level(Leq) may be calculated and this value should not exceed 90 dB(A).

Non-continuous exposure which cannot be adequately measured

In certain circumstances, for example where employed persons move from one area to another, it may be difficult to measure and control noise exposure to non-continuous sound. If the non-continuous exposure cannot be adequately measured and controlled, any exposure at a sound level of 90 dB(A) or more should be regarded as exceeding the accepted limit and requiring the use of ear protectors. Places where this level is likely to be exceeded should be clearly identified.

Sampling period for measurement of sound level

When making measurements to determine whether the acceptable limit is exceeded it will not normally be necessary to measure the sound level during the entire working period. Assessment may be based on sample periods which are typical of the working day.

Overriding limits

The A-weighted sound levels set out above are subject to an overriding condition that the unprotected ear should not be exposed to a sound pressure level, measured with an instrument set to the 'fast' response, exceeding 135 dB, or in the case of impulse noise an instantane -ous sound pressure exceeding 150 dB.
Other parts of the body should not be exposed to a sound pressure level, measured with the instrument set to the 'fast' response, exceeding 150 dB.

In 1973 the Industrial Injuries Advisory Council in accordance with the National Insurance (Industrial Injuries) Act 1965 reported on the question of whether there are degrees of hearing loss due to noise which satisfy the conditions for prescription under the Act (5). Amongst other things such a prescription demanded evidence, and the results from pure tone audiometry was one source. For example the degrees of disablement based on pure tone losses adopted by the American Academy of Ophthalmology and Otolaryngology gave the following scale; Hearing loss is averaged over 500, 1000 and 2000 Hertz (Hz) frequencies and disablement is taken as begening at 27 dB (ISO), by the addition of 1½ per cent for each decibel loss above 26 dB (ISO) the level of 100 per cent is reached at 93 dB. An alternative scale, also based on pure tone measurements, incorporating a higher starting point was also proposed. This is shown in Table 1.

Table 1

Hearing Loss	Percentage Disablement
1,2,3 kHz average (dB)	(%)
40	0
45	10
50	20
55	30
60	40
65	50
70	60
75	70
80	80
85	90
90	100
95	100 (or 110)

In Table 1, for purposes of calculating the average hearing loss of each ear, levels better than 0 dB or worse than 95 dB would be regarded as 0 and 95 respectively. Percentage figures above 100% would only be considered where deductions were made for presbyacusis; even then, any corrected figures still in excess of 100% would be regarded as no more than 100% for benefit purposes.

At the present time the limiting sound levels of the 1972 Code of Practice remain the standard in the U.K., and there is no requirement for routine audiometry. Some industries and the Ministry of Defence have well established hearing conservation programmes which include routine audiometry. King (6) gave a comprehensive account of the hearing conservation programme in the Royal Air Force and Bruton(7) has described a similar programme conducted by the Air Corporations Joint Medical Service at London airport. However because of variations in classification of criteria, definition of persons at risk, methods of audiometric measurement, interpretation of data particularly conversion to digital values in self-recording audiometers, periodicity of audiometry, storage and retrieval, comparison of such data can be difficult if not impossible. This does not imply criticism of the hearing conservation programmes of individual organisations but emhasises the need for uniformity if the data from such programmes are to be of value in any extended study. I would now like to consider the basic principles of hearing conservation, viz., noise exposure must be known and controlled , and each person's hearing must be measured before employment and at intervals throughout the period of employment if a noise hazard is thought to exist, in relation to the U.K. proposals.

Noise Exposure. The Industrial Health Advisory Committee's Noise Subcommittee, which produced the 1972 Code of Practice compiled a report titled Framing Noise Legislation, in 1975. The Working Party or Noise which was the successor to this subcommittee was formed to assist the Health and Safety Executive in the preparation of legislation recommended in the report. Considerable discussion has taken place on the current limiting sound level of 90 dB(A) for a steady state noise over 8 hours - equivalent values being apportioned for variations in level and duration. There is increasing pressure to review this with a lower level as the objective. There are three options; to leave the level at 90 dB(A), to reduce the level to 85 dB(A), to reduce the level to 80 dB(A). Taking practical and economical considerations it is my personal view that the final decision is likely to be 85 dB(A); this level is already adopted in some countries.

Having accepted a limiting level it is necessary to detail limits appropriate to the required criteria. It is important that these terms are clearly understood by those administering hearing conservation programmes. Guignard (8), explains this lucidly in Agardograph 151 in 1972, in relation to vibration; the explanation is equally valid for noise. A limit is regarded as a quantitative expression of an exposure level which could vary with the criterion adopted. This implies that the criterion of allowable exposure will be properly defined as an expression of the reason for limiting the exposure of a defined population.

Audiometry. The technique of measuring hearing by audiometry has been in existence for many years and has achieved various degrees of sophistication. For the reasons given earlier comparison of data from such measurements are not easy. Because of the need for uniformity among other things, the Health and Safety Executive of the United Kingdom circulated a discussion document titled 'Audiometry in Industry' in 1979, (9). We have agreed this document in principle. The scope of this document is for industrial audiometry, adapted to secure accurate quantitative information on the status of hearing before employment and during employment or both. It also gives guidance on indications for an audiometric programme, and on suitable techniques. I will summarise the relevant parts of this document;

- Audiometry means the determination of monaural hearing threshold levels for pure tones by air conduction. The pure tone frequencies should include 0.5, 1, 2, 3, 4, and 6kHz.

- The need for audiometry commences at 85 dB(A) Leq;

  i) Below 85 dB(A) Leq audiometry is not normally necessary.
  ii) Above 105 dB(A) Leq audiometry is mandatory.
  iii) Between 85 dB(A) Leq and 105 dB(A) Leq there is a corresponding increase in the desirability for audiometry.

  In practice it must be accepted that audiometry will be performed on all personnel working in a noise environment of 85 dB(A) Leq and above.

- A discrete frequency, pulsed-tone self-recording audiometer is recommended as the instrument of choice. Details of the recommended characteristics and calibration requirements are also given.

- Limiting noise levels within the audiometric booth are also given.

- Audiometric testing is carried out on a recommeded sequence as shown in Table 2.

Table 2

Test Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Left ear	0.5	1	2	3	4	6						0.5	1		2	3	4	6
Right ear							0.5	1	2		3	4	6					

. The audiograms are categorized on the basis of the sum of the hearing level for each
ear at the low frequencies 0.5, 1, 2,kHz and the high frequencies 3, 4, 6 kHz.

. If the sum of the hearing levels, either for low or high frequencies, shows an increase
of 30 dB or more when compared with the immediately preceeding audiometric examination,
or 45 dB when the interval of time since the preceeding examination exceeds three years,
the case should be categorized 1.

. If the difference of the sums of the hearing levels between the two ears exceeds the
following values the case should be categorized 2:

For low frequencies   45 dB

For high frequencies   60 dB.

. The sums of the hearing levels for each ear should be compared with the values in
Table 3 below, entering the table at the appropriate age. If the sum for either ear
exceeds the 'referral' level, either for low frequencies or high frequencies or both,
the case should be categorized 3. If the sum for either ear exceeds the 'warning' level,
either for low frequencies or high frequencies or both, but in no case exceeds the
'referral' level, the case should be categorized 4.

Table 3

| AGE IN YEARS | SUM OF HEARING LEVELS | | | |
| | 0.5, 1, 2 kHz | | 3, 4, 6 kHz | |
	WARNING LEVEL	REFERRAL LEVEL	WARNING LEVEL	REFERRAL LEVEL
20-24	45	60	45	75
25-29	45	66	45	87
30-34	45	72	45	99
35-39	48	78	54	111
40-44	51	84	60	123
45-49	54	90	66	135
50-54	57	90	75	144
55-59	60	90	87	144
60-64	65	90	100	144
65-	70	90	115	144

. Cases which do not fall into any of the above classes should be caregorized 5.

Storage and retrieval of data from such examinations can be a problem . We have designed a programme
which enables the data to be input to a microcomputer for direct categorization with a printout on
demand by the designated medical officer. Where self recording audiometers provide only an analog output
conversion to digital values for computer input should be carefully monitored and should be according
to the instructions given in the discussion document on audiometry. Currently we have linked a computer
audiometer to the micrcomputer which enables direct input from the print facility on the audiometer.
Storage is on disc and fast search, read &extend routines are incorporated.

I have attempted in the brief time to outline hearing conservatin with respect to the proposals in the
U.K. The use of audiometry as an essential tool to establish quantitatively the hearing status of an
individual and to monitor it during employment cannot be questioned. Audiometry also provides the
ultimate test of the success of any hearing conservation programme.

References:

(1)   Noise Final Report, Committee on the Problem of Noise. 1963. HMSO Cmnd 2056.

(2)   Proceedings of the International Congress on Noise as a Public Health Problem. 1973. Dubrovnik,
      Yugoslavia.   U.S. Environmental Protection Agency, Washington, D.C. 20460.

(3)   Ward, DW. Noise Induced Hearing Loss: Research Since 1973.   In; Noise as a Public Health Problem.
      Proceedings of the Third International Congress, 1978. Freiburg, W. Germany. The American Speech
      Language Hearing Association. 1980. Rockville Maryland.

(4)   Code of practice for Reducing the Exposure of Employed Persons to noise, 1972. HMSO

(5)   Occupational Deafness. Report by the Industrial Injuries Advisory Council, 1973. HMSO Cmnd 5461.

(6)   King, PF. Hearing Conservation in Aircrew and Ground Support Personnel. In Agardograph AGARD-AG-
      151, 1972.

(7)   Bruton, DM. The Administering of a Hearing Conservation Programme. In; ed. Robinson, DW.
      Occupational Hearing Loss. 1971. Academic Press, London and New York.

(8)   Guignard, JC. Vibration In Agardograph AGARD-AG-151, 1972.

(9)   Audiometry in Industry. Health and Safety Executive, 1979. HMSO.

DISCUSSION

DR. K.E. MONEY (CANADA)

Three speakers have now mentioned 85 dBA as the threshold for hearing damage. Could you comment on
the work of E. Shaw that suggests that in fact, if the noise exposure is 8 hours per day for a
working lifetime of 40 years, serious permanent damage can occur from exposure to noise even below
80 dBA and the limit to prevent damage should be 75 dBA.

DR. S. KANAGASABY (UK)

I agree that the work you cited, and other work, shows that damage occurs at 85 dBA, and I would
like to see a limit of 80 dBA, but 85 dBA is economically practical whereas 80 dBA is not.

FLYING HELMET ATTENUATION, AND THE MEASUREMENT, WITH PARTICULAR REFERENCE TO THE Mk 4 HELMET

G. M. Rood

Human Engineering Division
Royal Aircraft Establishment
Farnborough, Hampshire, England

SUMMARY

This paper is concerned with the effects of flying helmet attenuation upon communication performance, since it is becoming increasingly obvious that the flying helmet is providing the major share of the reduction of cabin noise to acceptable levels.

To predict the intelligibility of communication systems, it is necessary to be able to measure helmet attenuation accurately and repeatably, and it is this particular aspect which is highlighted. Hence, this paper discusses some of the results from a comprehensive series of tests involving subjective and semi-objective measurement of the attenuation of noise by flying helmets. The analysis shows that the semi-objective method of ascertaining hearing protector or flying helmet attenuation, using miniature measuring microphones, is a viable alternative to the existing standard REAT methods, and has considerable advantages in providing more useful information in less time. Additionally, high correlations exist between laboratory and in-flight measurements of attenuation, clearly indicating that laboratory measurements reproduce helmet attenuation actually found in the air.

THE SIGNIFICANCE OF HELMET ATTENUATION OF COMMUNICATION PERFORMANCE

Ideally a flying helmet would be worn by aircrew to provide crash protection, and to be additionally only a device onto which to hang an oxygen mask, a visor system, and a transducer to provide communication. Under these idealistic conditions without noise, or vibration, the quality of communications would rest only upon the quality of the communications system. Unfortunately, in current generation aircraft required to operate in the European theatre in the high-speed low-level role, and in some cases aircraft operating at higher altitudes, the cabin environment is not ideal and, of particular importance to communications, cabin noise levels in the region of 110-120 dB SPL are not uncommon.

To provide noise reduction at the source of this problem is difficult due to allocation of priorities, and even when some noise control is possible, the penalties of weight, cost and timescales conspire to make noise control at source a rapidly decreasing option when the possibilities of reduced fuel or weapon loads, or less avionics equipment, due to the extra weight of noise control measures, is considered. Thus, in spite of the excellent work in the noise control field, cost-effectiveness is low.

On the other hand, the cost-effectiveness of a flying helmet to provide noise protection, in addition to the primary task of impact protection, is high.

To further compound the problem, a survey of fixed and rotary wing military cabin-noise suggests that, at best, the existing cabin noise levels will remain at current levels, and at the worst, they will continue to rise.

Thus it is left to the protective helmet, to provide a solution to reduction in cabin noise to such a level that communication is not only possible, but efficient.

Whilst it is not totally impossible to communicate in noise levels in the region of 110 dB, restrictions on transducer power output, or in the case of the human, the restrictions due to the risk of hearing damage, make communications in levels such as this highly inefficient.

Thus unless the helmet reduces noise to a reasonable level at the ear then efficient communications will not be possible.

A brief example of this may be seen by taking the extremes of helmet attenuation to cover 95% of the aircrew population (ie mean ±2 standard deviations) for the RAF Mk 4 helmet; and to relate this to problems which can arise in the communications field.

A calculation of the Articulation Index (AI), to provide an estimate of the level of intelligibility of a communications system, shows that for given levels of speech, helmet telephone response etc, in a typical fixed wing cabin noise field, the change in attenuation between the $2\frac{1}{2}$ percentile and the $97\frac{1}{2}$ percentile alters the AI by 0.31, from 0.95 to 0.64, which under realistic conditions may degrade the intelligibility of speech by 12% or so. Similarly for tasks which involve some form of detection in noise, particularly in helicopter noise fields where high levels of noise in discrete bands are prevalent, a small reduction in noise level by improvements in helmet attenuation can provide large dividends in the reduction of noise masked thresholds.

However, provision of the required levels of helmet attenuation becomes difficult due to the conflicting priorities of space, weight, cost, etc. But having provided a helmet with a given attenuation, it is important to be able to measure the levels of attenuation both accurately and repeatably, such that when changes are made in helmet or earshell construction, the corresponding changes in attenuation may be accurately measured.

The following text is divided into two sections, one dealing with a comparison of the methods of measurement of attenuation, and the second with the correlation of laboratory measurements with in-flight measurements under operational conditions - both of which are necessary to provide data for an accurate predictive model for communications performance.

1   METHODS OF THE MEASUREMENT OF FLIGHT HELMET ATTENUATION

1.1   Introduction

Standardised methods of measuring the acoustic attenuation of flying helmets or hearing protectors, either by the British Method (BS 5108: 1974)[1] or the equivalent American Method (ASA STD-1: 1975)[2], use a subjective processing task in asking the subjects involved in testing to provide estimates of occluded and unoccluded threshold, with and without some form of hearing protector. The difference, in dB, between these thresholds is classified as the acoustic attenuation of the protecting device. Previous methods (ASA Z2422-1957)[3] had used a pure tone signal to provide the sound field and single frequencies of 125, 250, 500, 1000, 2000, 3000, 4000, 6000, 8000 Hz were used. The current standards use a refinement of this technique, and 10 one-third octave bands of random noise are used in place of the pure tone signals. This change of signal and a better noise field has provided a test with fewer error variances[4].

Whilst widely accepted, this real ear at threshold (REAT) method has several disadvantages, namely:

(i)   The test bands are generally on octave apart, and so no attenuation information is obtained for frequencies between the bands. The lack of this information could lead to mistaken conclusions in cases where a seal, earshell or cavity resonance occurs and causes a significant decrease in attenuation at a frequency not in any test band.

(ii)   It has been shown[5] that in using subjects to detect threshold values, masking of the lower frequency test signals occurs in the ear canal, thus providing an overestimate of hearing protection at low frequencies.

(iii)   The REAT methods preclude the testing of hearing protectors either in high noise environments or in the type of noise field in which the protector is to be used. This is particularly important in trials concerned with measurements in-flight, and is a technique which is currently being used in the measurement of active noise reduction reported in Paper 28. In addition, if a protector is tested by threshold methods, and is to be used in high noise environments, as is inevitably the case, the protector attenuation should be shown to be independent of noise level - which involves more testing and generally more assumptions.

(iv)   Subjective methods are both costly in capital equipment and time-consuming in testing, and are not particularly suited to type-testing or production testing.

Whilst these REAT methods provide valuable information and have been the mainstay of attenuation measurement over the last few decades, the methods have evolved historically from techniques that were possible at the time[6,7], when no reliable aternative methods were available. The relatively recent evolution of reliable miniature microphones with acceptable bandwidths has allowed techniques to be developed which, whilst retaining a few disadvantages, generally removes the constraints of threshold methods, and provides a considerably more flexibility of measurement.

Development of a technique using direct measurement of the noise field at the ear by the use of miniature microphones has been pursued at RAE in an attempt to provide a method which overcomes the constraints of REAT methods and provides equally, or more, accurate results, but retains the virtues of simplicity and cheapness whilst significantly reducing testing times. The method described was developed from techniques used by RAE for in-flight recordings in strike and other high-speed aircraft, which previously had provided analysis of communication system performance in the air.

Very simply, the method makes use of miniature or sub-miniature microphones which are placed under the flying helmet or protector earshell, affixed either to the subject at the entrance of the ear canal or to the earmuff opposite the ear canal entrance. A direct analysis is then made of the measured noise in the bandwidth required.

The role of the human as an accurate dummy rather than an intelligent participant has many advantages, and this technique of measurement is denoted as 'semi-objective' as distinct from 'subjective', where some intelligent processing is taking place, as found in the REAT techniques. The removal of the human element completely, by the use of an artificial head or ear, or an ear-model, is denoted as an 'objective' technique.

It is quite obvious that with all these types of testing, there will be some differences in the measured attenuation of a single protector caused by differences in methods; for instance on most artificial ears or heads air leakage from inside the earshell is not a problem, whilst it is a significant problem when using a human for testing. Similarly in the subjective REAT method neural or physiological noise is a problem, whilst it is negligible in semi-objective and absent in objective testing.

Two methods of measurement were used, namely subjective (to BS 5108: 1974) and semi-objective, and a comparison of the results indicate that a reliable and repeatable semi-objective method with all of its attendant advantages, may be used in place of REAT methods, one of the major advantages being that fewer subject-runs are required to maintain the accuracy of measurement relative to the REAT techniques.

1.2   Equipment

The hearing protectors

The three studies involved the use of two standard flying helmets, and one standard flying helmet with experimental earshells. The helmets comprised:

(i)   a standard in-service current RAF flying-helmet, the Mk 2/3 series helmet, which contains ear-muffs having seals of the low-compliance liquid-filled type;

(ii)   a standard Mk 4 flying-helmet, which is the service replacement helmet for the Mk 2/3 type. The helmet itself is of similar construction to the Mk 2/3, but the earshells are similar to those of the hearing protector and the seals are of the foam-filled, high-compliance type;

(iii)   a standard Mk 4 flying-helmet as in (ii) above, but with an experimental earshell of different dimensions and constructed of DMC.

1.3   Measuring microphone

The microphone type used throughout this series of semi-objective tests was a miniature electret microphone with dimensions of 8 mm × 5 mm × 2.25 mm, and a free field response of ±2 dB from 100 Hz to 5 kHz. The circuitry used, is such that levels may exceed 130 dB before noticeable distortion appears and was originally designed for use with a noise dosemeter to permit direct measurement of noise dose under the flying-helmet in flight[9].

Under laboratory conditions the microphone is placed on the subject's ear at the entrance to the external auditory meatus, using a flat lead (0.25 mm thick) to provide a signal path, this lead allowing minimal extra leakage past the helmet ear seals. In flight measurements, the microphone is fixed inside the helmet ear bun, opposite the entrance to the auditory canal.

1.4   Noise-generating and measuring equipment

The equipment was set up using the BS 5108 REAT rig at the Institute of Sound and Vibration Research (ISVR) at the University of Southampton[10]. This allowed the 10 one-third octave bands of random noise required for BS 5108 REAT tests to be produced. Non-coherent broadband noise was produced from four random-noise generators feeding through suitable spectrum-shapers to the same rig.

Analysis was by Bruel and Kjaer one-third octave real time analyser (Type 3347 and 2131). Bruel and Kjaer Type 2608 or 2609 low noise amplifiers were used to amplify the microphone signals where necessary.

The experimental layout is shown in Fig 1.

1.5   Subjects

For the three helmet studies, a specific population resembling aircrew was required. Fifteen students from the University Air Squadron were used and were considered to be a random selection of a specific population. This requirement was formulated since service aircrew generally have short hair and are clean shaven. By contrast many of the student population have long hair and/or beards which promote leakage past the ear shell seals, producing poor attenuation at low frequencies. All the Air Squadron students were currently in flight-training, or had flown regularly, and were used to wearing flying-helmets.

Anthropometric data in the form of head measurements were taken on all the flight-helmet subjects to allow a comparison with a wider ranging anthropometric study of aircrew[11,12].

1.6   Method of test

The methods described in this paper are the subjective REAT method and the semi-objective method using miniature microphone.

The REAT method allows investigation of the insertion loss technique alone to estimate the acoustic attenuation of the protector whilst the use of miniature microphones

allows both insertion loss and transmission loss to be measured. In this context inser-
tion loss is where the unoccluded ear is exposed to the noise field or band of noise,
and the threshold is determined or the noise level is measured; the ear is then occluded
with the protector and the measurements are repeated. The difference between the two
sets of measurements (in dB) is called the acoustic attenuation. Transmission loss is
where the protector is fitted and the noise at the ear and the noise external to the ear
shell are determined simultaneously, the difference in dB again being classified as the
acoustic attenuation.

The sequence of tests for the subjective/semi-objective comparisons is listed
below. It was arranged so as to provide a minimisation of error variance (eg the helmet
was not refitted between insertion loss and transmission loss measurements).

(1) Pre-test calibrations and helmet/protector fitting
(2) REAT randomised procedure: occluded/unoccluded
(3) Fit microphones to both ears
(4) Measure unoccluded: 10 one-third octave bands
(5) Measure unoccluded: broadband spectrum
(6) Don helmet
(7) Measure occluded: broadband spectrum
(8) Measure occluded: 10 one-third octave bands
(9) Fit outside microphone for transmission loss measurements
(10) Measure inside and outside microphones: 10 one-third octave bands
(11) Measure inside and outside microphones: broadband spectrum
(12) Completion of experiment. Post test calibrations.

1.7   Noise field

The noise fields were either the 10 one-third octave bands of random noise, or a
broadband approximation to pink noise over the range 50 Hz to 10 kHz.

The diffusivity of the field wad determined by measurement of the field of 27
points in a 300mm cube, the centroid of the cube being that position where the head of
the subject was placed during the tests. In the 24 one-third octave bands (50 Hz to
10 kHz), the maximum difference in any one band was 4.7 dB (80 Hz) whilst in the 100 Hz
to 5 kHz bands, the differences between any of the 27 points did not exceed 1.7 dB.

1.8   Results and discussion

The results presented here represent an analysis of the subjective versus semi-
objective tests and this represents the primary objective of the series of experiments.
Later analysis will further compare these two methods with those objective methods also
tested during the experiments (ie artificial ear results) to allow a three method compari-
son to be made, with the ultimate aim of providing a wholly objective test which provides
the 'correct' attenuation characteristic of a hearing protector or flying helmet. The
normal statistical studies have been carried out (though the details are not given here)
on the experimental data - and it was found that the normality of distributions, and
homogeneity of the data etc, would allow parametric analysis techniques to be utilised.

1.8.1   Insertion loss versus transmission loss (semi-objective data)

As noted previously it is possible, when using semi-objective methods, to use either
the insertion loss or transmission loss technique to measure the attenuation of a protec-
tive device. At RAE, the insertion loss is generally used in the laboratory whilst trans-
mission loss techniques are used for obtaining in-flight data. Thus it becomes important
to define the differences, if any, between the techniques and understand why they occur.

The grand mean values (the mean of tests 1 and 2 and of both ears) for the Mk 4
helmet are plotted against each other in Fig 2. In this particular case there are no
statistically significant differences (p < 0.01) in the frequency bands up to and includ-
ing the 3.15 kHz band except in the 1 kHz band which has a difference of 2.8 dB. In the
remaining 5 bands from 4 kHz upwards there is a significant difference in all bands.
Similarly for the remaining two helmets, the trend is the same between the two techniques,
but there are significant differences in some bands.

Since the differences at low frequencies are so small (effectively 1 dB up to 500 Hz)
there is a possibility that the higher frequency effects are due to the short wavelengths
and the precise placing of the microphones on the subject's ear. It was believed that
any systematic error variance from placing of the microphones would equate to zero, as
with the random error variance. Perhaps this is not so, and experiments to investigate
the field under the muff are now in progress. The most probable source of error, however,
is a small but consistent systematic error variance that may occur in the relative cali-
bration between the inner and outer microphones when using the transmission loss tech-
niques. Calibration appears to be critical, as errors of up to 1 dB at medium frequencies
and up to 3 dB in the four highest bands (5 kHz to 10 kHz) may occur. These, although
small in practical terms, tend towards the larger side for statistical purposes and are
systematic errors which do not tend to zero. This possibly may be resolved by a study
of the different calibration techniques now being carried out.

Whilst the differences are important and reasons for the discrepancies must be explained,
the problems of differences between statistically significant and significant practical
differences arise. With the semi-objective data, 60 data points are processed for each

frequency and generally standard deviations fall below 5 dB - except at the highest frequency bands - so that the standard error is approaching 0.5 dB. With the lower frequency band standard deviation of less than 2 dB, the standard error approaches 0.35 dB. Thus differences between means of 1.5 dB can be statistically significant. Thus there is a discrepancy between a difference of *practical* significance, say 2 dB or more, and the corresponding *statistically* significant difference which may be 1 dB or less in some cases. In these cases it is felt that whilst the differences may be significant (and should be further analysed to find an explanation), in practice it would be permissible to conclude that, apart from the higher frequency bands, both techniques provide the same answer. This is more clearly seen in Fig 3 where the differences for all frequency bands is shown for all three helmets. Up to the 4 kHz band the differences between the two techniques is less than 3 dB, with 10 bands out of 20 having differences less than 1 dB for the Mk 2/3 and 16 out of 20 for the Mk 4 helmet. Above 4 kHz the differences become greater, rising to a maximum of 5.8 dB in one case in the 5 kHz band.

Generally it is felt, however, that either technique may be used, and the differences up to the 4 kHz band do not exclude either technique being compared with the other reasonable confidence, which will however, be quantified as further research into the differences progresses.

1.8.2 Replications: runs 1 and 2

REAT methods use a single replication since analysis of REAT data has shown that the intra-subject variance is small compared with inter-subject variance and thus more subjects are more important than more replications. The results obtained here support this statement since for the REAT data obtained for all three helmets, there is no significant difference in any of the 10 frequencies between replication. Similarly for the semi-objective tests, the results for the grand mean of left and right ear for the Mk 4 helmet by the insertion loss method and for the Mk 2/3 helmet for the transmission show there is an insignificant difference between runs. For all three helmets there are only two bands where the significance of the differences between replications rises to 1%, in the Mk 4 alt helmet at 3.15 kHz for the insertion loss technique (0.76 dB), and for the same helmet in the 50 Hz band (0.85 dB) with the transmission loss technique.

The summary of Fig 4 of the comparison between replications for the insertion loss technique for all three helmets shows clearly that for all practical purposes there is no difference between runs and that to perform only one run would entail no loss of accuracy.

1.8.3 Differences between left and right ears

With the use of miniature microphones, sound pressure levels in both ears may be measured, and, unlike the REAT method, differences between ears may be shown. The mean values for left and right ears for both runs were combined. The plots in Figs 5 and 6 show samples of the helmets and highlight the differences which occur between ears. For example, in Fig 5 whilst there are only significant differences (p < 0.01) in the 1 kHz band, and all frequency bands above and including 5 kHz; in Fig 6 differences of the same significance are apparent in no fewer than 16 of the 24 bands. The reasons for these differences are under investigation and no reasons are yet apparent, but it is interesting to note that when left and right ear data are combined to form a mean for separate runs, no significant differences occur. The plot in Fig 7 shows a correlation between left and right ears for the Mk 4 helmet. Thus, in more general terms it can be seen that there is a highly significant correlation (r = 0.993) between left and right ears and equation of the line, y = 1.069x - 1.134, when compared with theoretical y = x shows an insignificant difference p to the 1% level of either slope or intercept. Thus, in the first instance, setting aside the individual differences, left and right ear attenuations are highly correlated. However, until rational explanations are available for the individual differences, the mean of both ears should be used where possible.

1.8.4 Differences between broadband and sequential band results

When REAT methods are utilised, data is only available for the 10 one-third octave bands of noise used and information between these bands is lost. The use of broadband noise allows information over a wide band to be obtained.

Fig 8 compares the attenuation results using broadband noise and those obtained with the 10 one-third octave bands under similar circumstances. The greatest difference for this helmet is less than 1 dB and for all helmets is less than 2 dB except for the Mk 2/3 helmet in the 8 kHz band where the difference is 2.6 dB. When summarised, the differences between attenuation measured from one-third octaves and those measured by broadband noise are insignificant, and thus broadband noise may be used and this permits essentially three times as much data to be obtained in significantly less time.

1.8.5 Differences between REAT and semi-objective results

The final graphs, Figs 9 and 10, exemplify comparisons between subjective (REAT) and semi-objective insertion loss tests for the three flying helmets. Each protecting device was measured using the same subjects for each technique, and, as detailed previously, during the same session.

Inspection of the figures shows three noticeable trends which were apparent with all devices tested. At low frequencies, that is, in the 63, 125 and 250 Hz bands, the attenuation measured by the REAT method significantly exceeds that measured by the comparative

semi-objective method. A two-tailed 't' test shows significant differences (p < 0.01) in the 63, 125 and 250 Hz bands. At medium frequencies between the 500 Hz band and the 4 kHz bands there is a good agreement between the two methods, whilst above this frequency there is a divergence with significant differences, again with the REAT values exceeding the semi-objective values in all four cases.

Fig 11 illustrates a possible shortcoming of the method, and uses data on a standard hearing defender to illustrate the problem. An interesting dip in the REAT results is shown in the 2 kHz band, suggesting perhaps a bone conduction threshold limit, which, of course, would not be shown by the microphone method. When these types of body/bone conduction thresholds are reached, one of the limitations of semi-objective methods becomes apparent, and research is needed to find reliable body/bone conduction thresholds to provide corrections for semi-objective methods. This dip does not show in the helmet results, probably because the outer shell of the helmet attenuates the higher frequency noise to below the possible bone conduction limit.

1.9 Conclusions

The results show that the use of a miniature-microphone semi-objective method may be used successfully in place of the standardised REAT methods. The standard deviations are not significantly different, and with the smaller standard errors of the mean and the amount of additional data available over the extended frequency range of testing, it is felt that the method provides an attractive, accurate and simpler alternative to REAT methods - methods which are perhaps anachronistic. It is clear that either the insertion loss or transmission loss technique may be used and the results using either agree will from the 50 Hz band to the 4 kHz band, (ie the mean attenuation values are all within a maximum of 3 dB of each other - in the majority of cases within 2 dB). In the four remaining bands above 4 kHz there is a divergence and differences up to 5 dB can occur. However, either insertion or transmission loss techniques may be used with confidence up to 4 kHz, and where frequencies above this are required, the insertion loss technique, with no danger of calibration errors, will provide reliable data. In general the insertion loss technique is used at RAE in the laboratory; but where high levels of noise are required (ie above an Leq of 75 dB(A) at the subject's ear) or in-flight measurements are necessary, only the transmission loss technique is possible.

A further area requiring investigation is whether or not there are discrepancies between the attenuation of the left and right earshells. Generally there is a very high correlation between left and right ears, but individual differences provide some scatter. However when both ears are summed and the mean value taken and compared between the replications then no significant difference occurs. Thus, providing the data at both ears are used, where possible, no loss of accuracy will occur. As part of this particular exercise it is also shown that there is no difference between mean values for replications, and thus a single run is adequate, and as noted previously this is also true for all the REAT tests, where neither the means nor variances are significantly different.

The conclusions to be drawn from the comparison of the broadband noise data and that obtained from the one-third octave data is that the differences are insignificant and that, quite apart from the obvious advantages of obtaining the additional information contained in the bands between the one-third octave measurement bands, either method may be used to measure the attenuation characteristic. The further major advantages are that all 10 or 24 frequency bands may be measured simultaneously instead of sequentially, providing a considerable saving in testing time, and possibly in accuracy, if the noise field is not essentially stationary. Additionally the broadband spectrum may be analysed, if necessary, in frequency bands narrower or broader than the one-third octave bands, allowing a greater flexibility in analysis, which will assume some importance if, for example, predictions of acoustic masking are required in noise fields which contain a high proportion of discrete tones, as in helicopter noise or in some armoured fighting vehicle noise.

The final comparison, that of the REAT method against the semi-objective method using miniature microphones poses a more fundamental question: that is, which is the 'correct' method of attenuation determination? Obviously over the medium frequencies from 500 Hz to 4 kHz there is no significant difference, and at the higher frequencies the differences may, at least in the first analysis, be looked upon as academic from a practical viewpoint, since 6 or 7 dB difference on an attenuation of over 40 dB is negligible. Nevertheless it may be important and research is in progress to try to explain the differences. The major problem, at least in the aircraft industry, is at low frequencies where there is a great deal of noise, and there is difficulty in increasing the attenuation of flying helmets due to the constraints of size and weight. One would wish to accept the markedly greater results obtained by the REAT method, but the research of Anderson and Whittle of the National Physical Laboratory[5], convincingly suggests that these figures are due to masking by physiological noise in the auditory meatus. If this is the case, the attenuation measured by the miniature microphone is likely to be closer to the 'correct' value than the REAT figures.

It is considered that valid results can be obtained by semi-objective methods using the transmission loss technique with measurements at both ears of 15 subjects in a single run. The major advantage would be that the testing time - using computer analysis techniques - would be reduced by around 95%. The testing time for a single subject by REAT methods is approximately 1 hour for two runs, whilst with the proposed semi-objective method the testing time can be reduced to 1 minute, plus a further 2 minutes or so for fitting and removal of the microphones. Thus it should be possible to carry out a single comprehensive test on a flying helmet or hearing protector, which will give data as

accurate as and probably more accurate than, those from the corresponding REAT tests, and significantly reduce the overall testing period.

2   MEASUREMENT OF HELMET ATTENUATION IN-FLIGHT AND CORRELATION WITH LABORATORY MEASUREMENTS

2.1   Introduction

As a necessary extension to measurement of helmet attenuation in the laboratory a trial was undertaken to measure the attenuation of flying helmets in-flight, since it seems pointless to measure attenuation or changes in attenuation with changes in helmet or ear shell form, if these changes are not translated into practice in the air.

In this case, a series of attenuation measurements were made in-flight in Jaguar aircraft during operational sorties from RAF Coltishall. Pilots from 6 Squadron and 41 Squadron carried out a series of attenuation measurements on both their own Mk 2/3 series helmets and production Mk 4 helmets.

For the Mk 2/3 tests, 14 pilots provided five attenuation measurements during each flight, although in a few cases this was reduced to three replications due to operational flying requirements. Similar tests were carried out with the Mk 4 helmet, in this case 13 pilots acting as experimental subjects.

2.2   Experimental procedure

In each case, the noise field was measured at the ear under the ear shell and external to the helmet using the transmission loss technique.

Before each flight a miniature microphone was affixed inside an ear shell opposite the ear canal entrance with care being taken that the microphone was not in a position to be occluded by the pinna. The outer microphone was fixed to the surface of the helmet in a position opposite to the inner microphone.

The pilot was fitted with he equipment prior to entering the aircraft. The control box for turning the tape recorders on, and telephone line off, was strapped to the pilot's left thigh and the two Nagra SNN recorders were placed in each leg pocket.

The procedure during the flight for the attenuation measurement was as follows:

At a period during the flight, the pilot turned on the tape recorders and announced the aircraft speed and height. The input to the telephone line was then disconnected for a period of 30 seconds or greater, the pilot keeping his speed and height constant, and keeping his head still. During this period the only noise reaching the inner microphone was from cabin noise. After this period, the telephone line was re-connected and the end of the first trial announced. This was repeated a further four times during the flight.

Synchronisation between the two tape recorders was achieved by a switching system which was such that as the telephone system was disconnected by the switch on the control box, the noise fields of the outer and inner microphones were switched between tape recorders allowing accurate synchronisation to be carried out.

2.3   Calibrations

In each flight case, the complete system was calibrated, firstly by the use of a 94 dB 1 kHz pure tone from a B and K calibrator, thus setting the absolute levels, and secondly by calibrating each microphone against the other by calibrating in a broadband noise field (50 Hz to 10 kHz), thus setting relative levels. In addition, calibrations were made, in broadband noise, of each microphone and recording system against a standard half inch B and K microphone (type 4134) and cathode follower (type 2619), to allow corrections to be made for the frequency response of measuring microphones. All calibrations were measured on a B and K type 3347 digital one-third octave real time analyser.

2.4   Analysis

Analysis was carried out immediately post-flight, using the same type recorder as in-flight and was in one-third octave bands using a B and K 3347 real time analyser controlled through a Hewlett Packard HP 9825A mini-computer system. The analysis was performed generally over a 32 second period, with four averages being taken in this time.

In these analyses great care was taken to ensure that noise-floor and/or dynamic range problems did not contaminate the results, since this can be a potential problem with this type of in-flight measurement.

2.5   Results and discussion

Very simply, the results can be reduced basically to two plots, shown in Fig 12 and Fig 13, which represent the results of the grand mean attenuation of the Mk 4 and Mk 2/3 helmets measured in-flight against attenuation measured in the corresponding laboratory tests, described earlier.

In the case of the in-flight work it was only possible to measure the attenuation of the helmet at one ear, since restrictions upon the number of tape recorders the pilot

could carry were necessary, and the Nagra miniature stereo recorders were not developed to the extent required for this trial. In the laboratory, measurements were made at both ears.

The plots for both helmets generally show a good correlation between flight and laboratory measurements. In the Mk 2/3 data, Fig 13, there are statistically significant differences in some frequency bands, the greatest differences being around 5 dB in 4 frequency bands, with similar results being shown in Fig 12 for the Mk 4 helmet data. The question now arises again on the difference between statistically significant differences of results and differences which have a practical significance.

In measurement of helmet attenuation, especially outside of the laboratory and with the less stringent control of the variables contributing to the attenuation measurement, it is considered that 2 dB, which in many cases represents a jnd (just noticeable difference) in intensity for the type of noise, is a reasonable figure to regard as a practical minimum difference in intensity. Thus it can be seen that in the case of the Mk 2/3 helmet, approximately 63% of the differences between flight and laboratory fall under the 2 dB levels, with 71% of results falling below 3 dB. The corresponding figure for the Mk 4 helmet being 58% at the 2 dB level and 79% at 3 dB.

This difference may be explained by the slight differences in the total number of measurements recorded. In the laboratory both ears were used, whilst the in-flight measurements were available at only one ear. This difference between left and right ears explained previously may account for some of these differences. As an indication, a comparison was made between the in-flight data for both helmets against the left and right ear data separately. In the case of the Mk 2/3 helmet for the left ear approximately 50% were less than 2 dB in difference, with 29% for the right ear, the 3 dB figure being 58% and 54% respectively. Corresponding figures for the Mk 4 helmet were 63% and 50% for the left and right ear respectively, the 3 dB figure being 88% and 67% respectively.

The indications are that, in general, differences between laboratory and in-flight measured attenuation are small and it is possible to show that within the limits of individual variance, that no significant difference exists between flight and laboratory measurement by plotting the in-flight data against laboratory data.

Fig 14 shows such a linear regression plot of the laboratory data against flight data for the Mk 2/3 and clearly shows a highly significant ($p < 0.001$) correlation coefficient ($r = 0.989$) and an equation of the line, $y = 1.07x + 0.13$, which is insignificantly different from the theoretical $y = x$ in both slope and intercept. Similarly for the Mk 4 helmet, Fig 15, the correlation of the regression ($r = 0.989$) is highly significant ($p < 0.001$) and the equation of the line, $y = 0.99x + 0.82$, shows again, an insignificant difference in both slope and intercept. Thus, in general, leaving aside individual differences in frequency bands, a highly significant correlation exists between flight and laboratory measurements of helmet attenuation.

However, there are individual differences in the attenuation measurements and work is in progress to attempt to explain the discrepancies and to quantify the accuracy of the methods.

2.6   Conclusions

(1)   Using the techniques described to measure flying helmet (or hearing protector) attenuation, there is a highly significant correlation between laboratory and in-flight measurements.

(2)   The technique of semi-objective attenuation measurements using miniature microphones provides reliable and repeatable data and has the following advantages over standardised REAT techniques.

(a)   It provides up to three times more useful data in essentially 10% of the time.

(b)   The data may be analysed in any bandwidth required (*ie* octave, one-third octave or narrow band).

(c)   Equipment may be tested at noise levels in which the equipment will be used, and is a technique which is readily adaptable for measurements of attenuation in the field.

(d)   There is a very high probability that the technique will provide a more 'correct' result than REAT methods at low frequencies.

(e)   The capital and running costs are significantly reduced.

REFERENCES

1 British Standards Institution. British Standard method for the measurement of real-ear attenuation at threshold of hearing protection BS 5108 (1974)

2 Acoustical Society of American Standard. Method for the measurement of real-ear protection of hearing protectors and physical attenuation of earmuffs. ASA STD 1-1975

3 American Standards Association. Measurement of the real ear attenuation of ear protectors at threshold. ASA Z24.22 (1957)

4 L.S. Whittle, D.W. Robinson. On the measurement of real-ear attenuation of hearing protectors by standardised test methods. National Physical Laboratory Acoustics Report Ac 79 (1977)

5 C.M.B. Anderson, L.S. Whittle. Physiological noise and the missing 6 dB. Acustica Vol 24, pp 261-272 (1971)

6 S.S. Stevens, D.R. Griffin, S.J. Goffard, J. Volkman. The acoustic design of earphone sockets helmets and headsets. OSRD Report Contract 1366 (1942)

7 D.R. Wilkie. The acoustic properties of flying helmets FPRC 732 (1950)

8 H.C. Attwood, K.R. Maslen. Noise attenuation of Mk 2 and Mk 3 integral helmets - a survey of results from different sources 1964-1970. RAE Technical Memorandum EP 453 (1970)

9 G.M. Rood, D.C. Baines. A meter for the measurement of noise dose under circumaural hearing protectors. RAE Technical Memorandum FS 96 (1976)

10 E.M. Whitham, A.M. Martin. The ISVR facility for the evaluation of hearing protectors following the British Standard procedure. ISVR Memorandum 558 (1975)

11 C.B. Bolton, M. Kenward, R.E. Simpson, G.M. Turner. An anthropometric survey of 2000 Royal Air Force aircrew 1970/71. RAE Technical Report 73083 (1973)

12 P.C. Hobbs. An anthropometric survey of 5000 Royal Air Force aircrew heads, 1972 RAE Technical Report 73137 (1973)
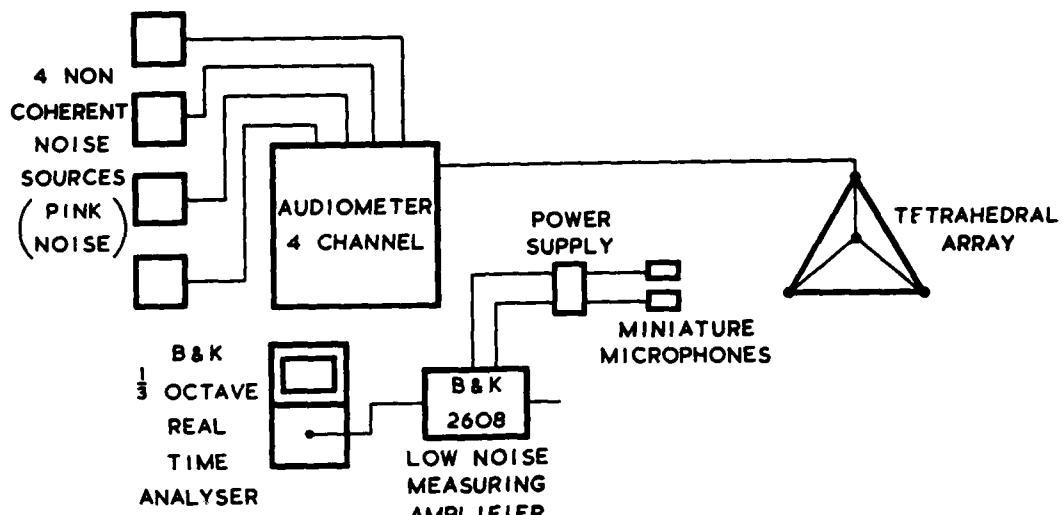
Fig 1   Test equipment arrangement
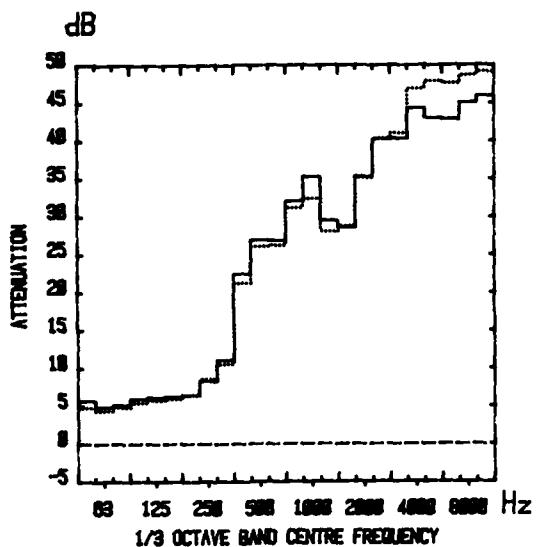
Fig 2   A comparison of insertion loss and transmission
loss data: Mk 4 helmet
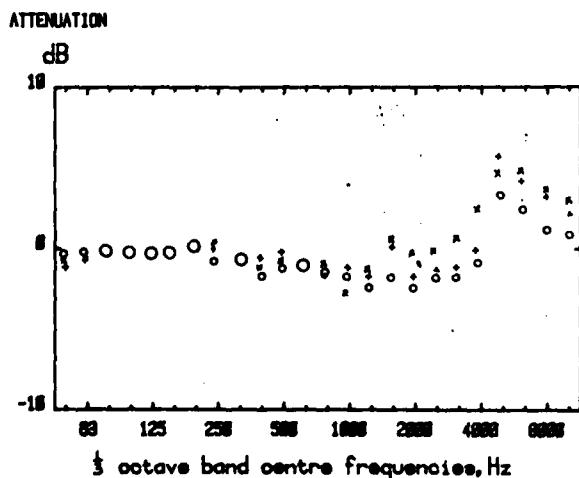
Fig 3   Summary of transmission loss versus
insertion loss data
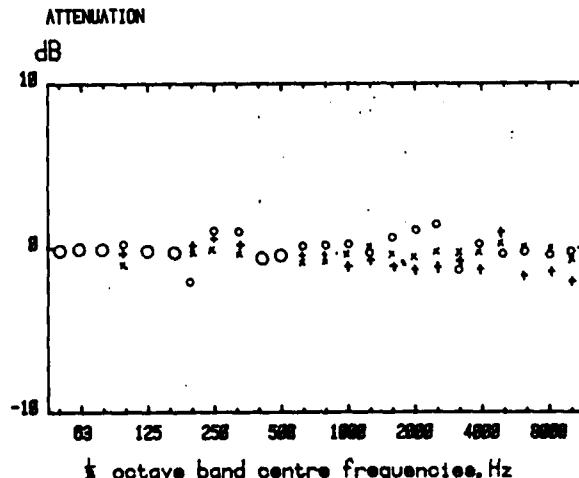
Fig 4   Summary of replication data:
Run 1 versus Run 2
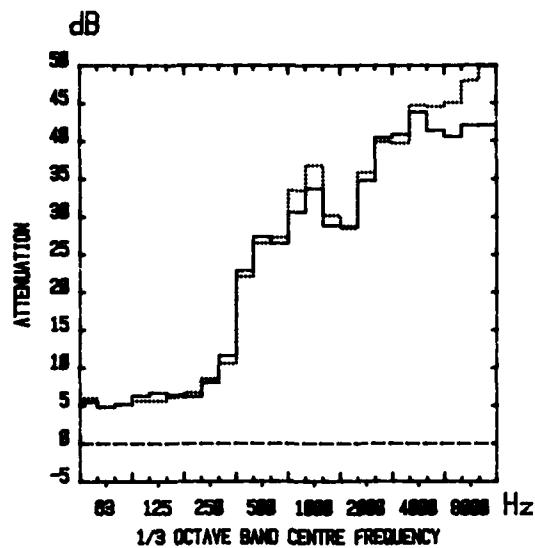
Fig 5 Left versus right ear data:
Mk 4 helmet: insertion loss
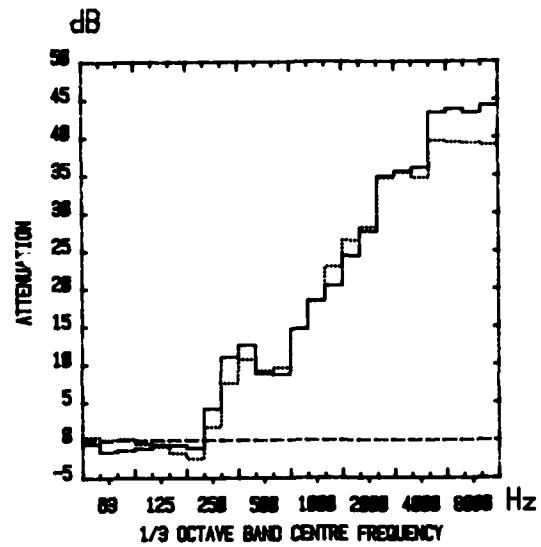
Fig 6 Left versus right ear data:
Mk 2/3 helmet: transmission loss

$r = 0.993$
$y = 1.069x - 1.134$
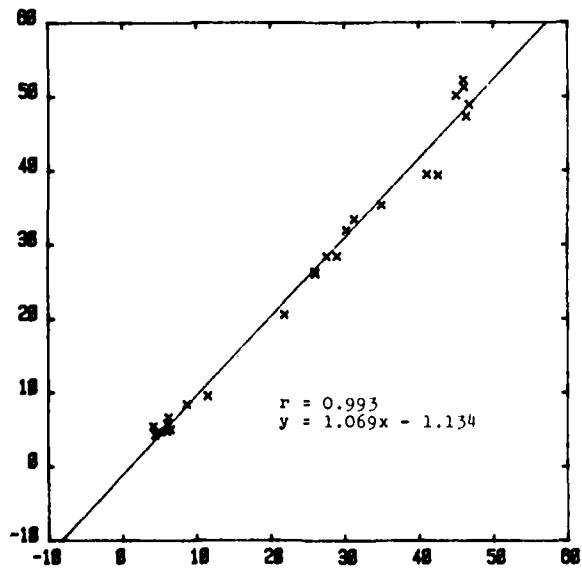
Fig 7 Correlation of left and right data:
Mk 4 helmet: transmission loss
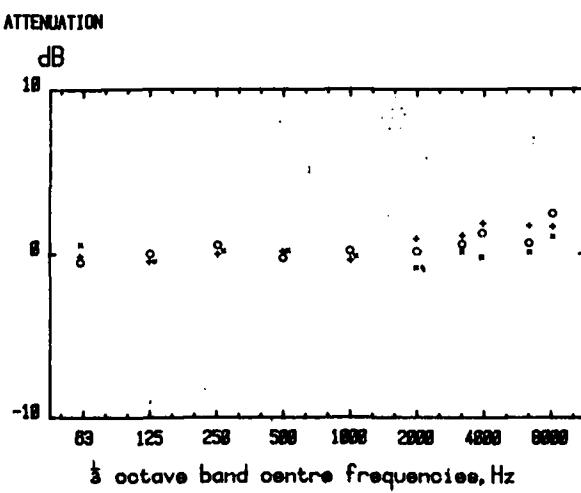
Fig 8 Summary of broad band versus one-third
octave band data
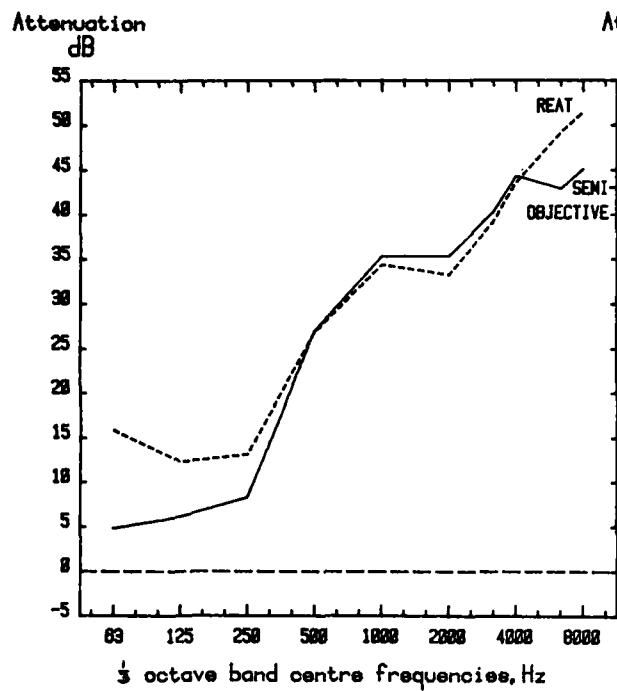
Fig 9   A comparison of REAT and semi-objective
        attenuation data: Mk 4 helmet
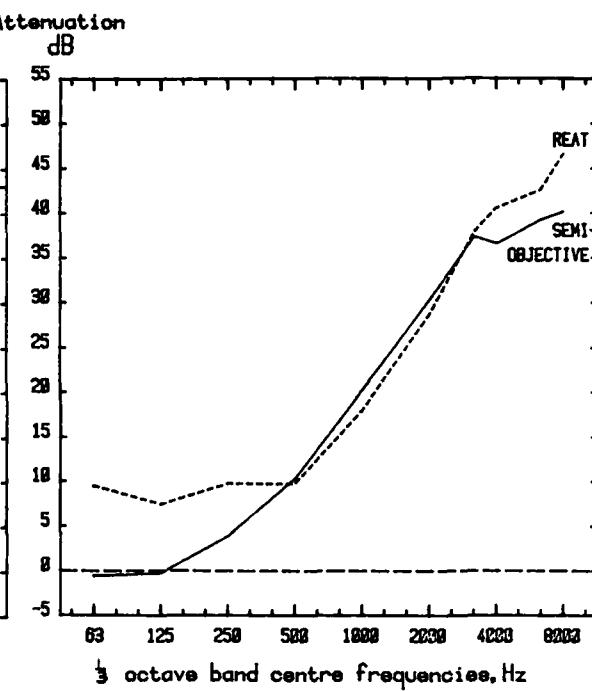

Fig 10   A comparison of REAT and semi-objective
   attenuation data: Mk 2/3 helmet
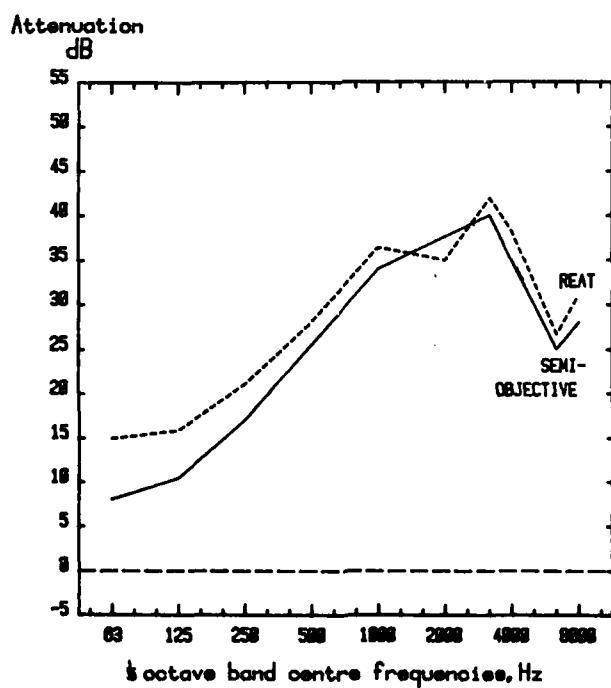
Fig 11   A comparison of REAT and semi-objective
   attenuation data: hearing protector
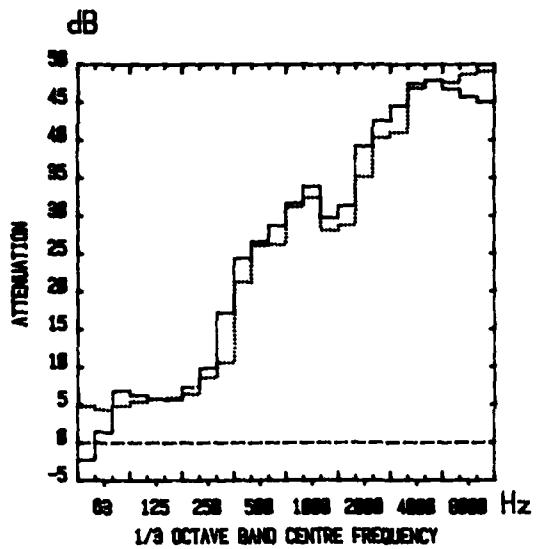
Fig 12 A comparison of laboratory and in-flight measured attenuation: Mk 4 helmet
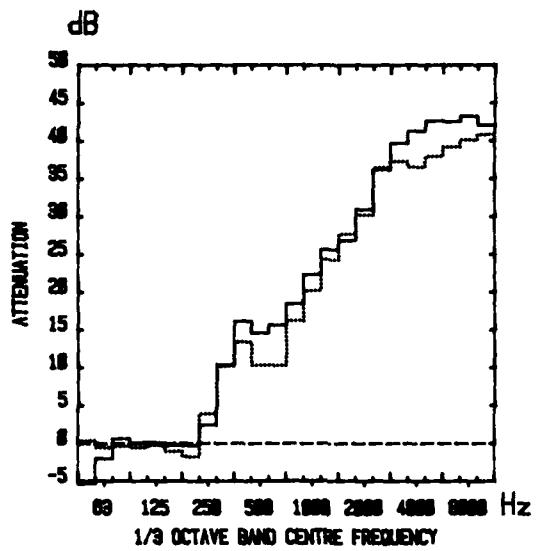

Fig 13 A comparison of laboratory and in-flight measured attenuation: Mk 2/3 helmet


r = 0.989
y = 1.07x + 0.13
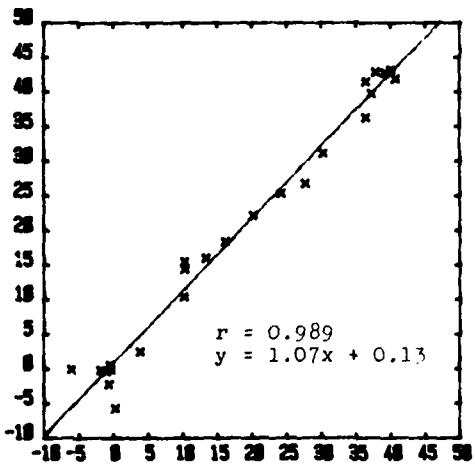
Fig 14 Correlation of laboratory and in-flight measured attenuation: Mk 2/3 helmet


r = 0.986
y = 0.99x + 0.82
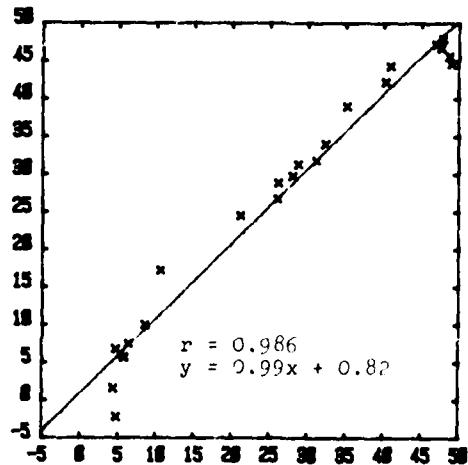
Fig 15 Correlation of laboratory and in-flight measured attenuation: Mk 4 helmet

A MULTIPLEXED DIGITAL VOICE
INTERCOMMUNICATIONS SYSTEM COMPATIBLE
WITH FUTURE VOICE CONVERSION TECHNIQUES

by

Richard F. Bolt

Barry D. Sanderson

Applied Research Department

U. S. Naval Avionics Center

6000 East 21st Street

Indianapolis, Indiana

46218

U. S. A.

SUMMARY

The U.S. Naval Air Systems Command's Avionics Components and Subsystems (AVCS) Program is developing a modular multiplexed digital voice system designed to counter Fleet Intercommunications System (ICS) deficiencies. This system will be usable in a variety of aircraft in the U.S. Navy, Army and Air Force. The individual stations will be interchangeable from aircraft-to-aircraft, and the stations will be comprised of plug-in modules that can be easily replaced by maintenance personnel. This modularity will allow the system to be updated with future voice conversion techniques without rework or redesign. Currently available voice conversion techniques have been evaluated for intelligibility in some military acoustic noise environments.

The ICS Built-In-Test will automatically locate faults and identify failed modules. This use of Built-In-Test will reduce maintenance costs significantly. The use of state-of-the-art devices will enhance the performance and availability of the system. The application of modern digital design practices to this ICS will reduce or eliminate the problems of crosstalk, Electro-Magnetic Interference (EMI), and lack of flexibility in station selection and system configuration.

The immediate tasks in the AVCS ICS development are to complete the voice digitization selection and continue package configuration investigations. Tri-Service coordination will be maintained throughout the specification preparation and full scale development.

INTRODUCTION

The aircraft ICS provides audio communications among crew members, radios, and mission related equipment. Traditionally each new aircraft has been accompanied by the development of a new ICS or an extensive modification of an existing system. The existing systems are all-analog and were developed without commonality considerations. Six types of ICS systems are currently used on U.S. Navy/Marine aircraft. Under the AVCS Program, the U.S. Navy plans to develop a standard digital voice ICS that is compatible with MIL-STD-1553B, meets EMI requirements, and provides reduced size, weight and power consumption.

In recent years, various problems have been identified in Fleet ICS systems. These problems can be broadly grouped into two areas: performance and equipment availability. Performance problems consist of poor isolation between communication channels, poor isolation from EMI sources, and a lack of flexibility in station selection and system configuration. Equipment availability problems result from the use of obsolescent technology, proliferation of equipment types, high skill requirements of maintenance personnel, and high cost of logistics support due to a lack of commonality between systems.

State-of-the-art technology can be used to improve both performance and equipment availability of Fleet ICS. The key to improving ICS performance is the application of digital processing to the audio information. By digitally encoding the audio signal at the ICS station and transmitting the information as time-division multiplexed data, the encoded signal can be isolated from signals from other ICS stations. This technique also provides isolation of the encoded signals from EMI noise sources. Signals from all ICS stations are available in digital form, allowing increased flexibility in station selection and system configuration to be achieved. System

availability improvements will result from the proper system architecture and the increased capabilities of the electronic devices within that architecture. The implementation of this architecture to interconnect intelligent ICS stations can provide several system advantages: graceful degradation, self-test, fault isolation, and module replacement maintenance.

In 1978, the Naval Air Systems Command initiated the Avionics Components and Subsystems Program. This program will plan and develop a family of standard avionics, supportive of, but separate from, major aircraft weapon system acquisitions and common to multiple aircraft. This common equipment must meet the performance requirements of all aircraft. The advantages of this equipment commonality are realized during its service life. Several benefits are: improved system configuration management, minimized training to operate and repair the equipment, and reduced logistics support costs. Performance and equipment availability problems in Fleet avionics will be reduced or eliminated by the development of new systems that meet the following design goals:

1. Increase performance by identifying and correcting design deficiencies in existing equipment

2. Reduce proliferation of equipment by developing common types of equipment for use in multiple aircraft

3. Reduce technology obsolescence by developing systems that are upward compatible with new technologies, and

4. Reduce maintenance by design for repairability and use of Built-In-Test (BIT) techniques.

These goals have governed the ICS development at the Naval Avionics Center under the Avionics Components and Subsystems Program.

The remainder of this paper consists of four major sections. The first section provides background information and identifies design issues in the development of the next generation ICS. The second section describes the system that the Naval Avionics Center has designed and built as a laboratory tool in developing the next generation ICS. The third section presents a solution to current ICS problems. The last section summarizes current ICS problems, restates the proposed solution to these problems, and recommends future efforts.

BACKGROUND

Some ICS deficiencies are inherently present in their system architecture. Existing ICS architectures are either distributed analog or central switch types. The distributed analog type uses parallel wires between ICS stations for interconnections. This wiring provides an easy way for EMI to enter the system and degrade its performance. Additionally, the flexibility of station selections is limited by this wiring. In order to allow a new set of station selections, additional wiring is required between stations. The central switch type requires less wiring between stations, however, microphone and headset wiring is routed from each station to the central switch. These microphone and headset wires provide an easy way for EMI to enter the central switch system. The station selection flexibility of the central switch system is limited by the growth potential designed into the central unit. If the central switch was designed for a particular aircraft, it probably lacks sufficient flexibility for use in other aircraft.

The Garrett Air Research Corporation, under a contract to the Naval Air Development Center, designed and built a system to demonstrate the feasibility of transmitting audio data in digital form via a MIL-STD-1553 data bus. The Garrett system used Continuously Variable Slope Delta (CVSD) modulation as the voice digitization technique. This system proved that encoded audio could be successfully transmitted via a MIL-STD-1553 data bus.

The design issues for the next generation of ICS may be divided into two categories. The first concerns the solution to current ICS problems. The second category deals with design issues raised by solutions to the current problems. Presently, some aircraft handle both secure and clear audio data with two separate ICS in order to maintain the necessary isolation. The next generation ICS should be designed to provide both secure and clear communications without the redundancy of two separate ICS. System commonality, across all aircraft considered, requires sufficient functional capacity to handle the maximum audio traffic and packages to fit available space.

Fleet equipment availability will be improved by the use of modularity in Intercommunication Systems. The next generation ICS needs to provide both electronic functional and mechanical modularity. Electronic functional modularity is needed for self-test for isolating system faults, while mechanical modularity provides easy field repair. An ICS design goal is to develop a system that tests itself and identifies the faulty functional modules to be replaced.

Two remaining design issues are raised by the solutions to current ICS problems. The first is the need to provide a system intelligibility at least equal to current systems. The second issue is the need to be upwardly compatible with future voice conversion techniques.

DESCRIPTION OF MUDIV SYSTEM

The Multiplexed Digital Voice (MUDIV) evaluation system was designed and built as a laboratory tool for the development of the next generation ICS. The MUDIV system consists of three ICS stations and the interconnecting MIL-STD-1553 bus. The MIL-STD-1553 bus consists of a single twisted shielded pair of wires. Each ICS station contains three 8 inch by 9.75 inch circuit boards. Each station consists of an enclosure containing the following: the circuit boards, power supplies, connectors, front panel controls (switches and potentiometers), and front panel indicators. The bus controller is on one circuit board housed in the Master station. With the exception of the bus controller board and one front panel switch for this Master station, all three ICS stations are identical. Thus, parts of one station can be interchanged with the corresponding parts of either of the other two stations.

The MUDIV ICS provides audio communication via selectable paths among the stations. With three stations, there are only two useful types of communication paths. One type is a private communication path between a pair of stations, excluding the third station. The other type is a "party line" which allows all three stations to communicate with each other. As the number of stations increases, the number of potentially useful paths increases rapidly. The basic design of the MUDIV system allows many of these useful paths to exist simultaneously, with the operators choosing which of these paths are used in any given situation. This capability eliminates systemic limitations on the flexibility in station selection.

The function of the bus controller is to issue commands to the ICS stations so that audio communications among the stations occur in the desired manner. The MUDIV is used in evaluating various voice digitization techniques; therefore one command issued by the bus controller informs each station of the voice digitization technique currently selected by the operator of the Master station. The voice digitization technique is selected by the mode selector switch on the Master station. The other commands issued by the bus controller direct each ICS station in turn to transmit audio data onto the bus and the other ICS stations to receive this audio data from the bus according to the currently active communication paths.

Since all three ICS stations perform the same functions, the operation of only one station will be described. Usually the three stations are not all performing the same function at any given time, but function together in a coordinated manner. For example, when one station is transmitting audio data onto the bus, the other two stations are (if so instructed) receiving this audio data from the bus. The three circuit boards within a station are named the Analog board, the Buffer board, and the 1553 board. These are diagrammed in Figure 1. The Analog board provides the primary man-machine interface for the station. This board is devoted to the digitization of incoming microphone signals, reconstruction of the audio output, and amplification of headset audio. Ten different bit-rate/audio bandwidth modes are available on the Analog board. In five of these modes the analog filters reject frequencies above 3500 Hz, while in the other five modes the analog filters reject frequencies above 4500 Hz. The two fundamental voice digitization techniques available in the MUDIV ICS are Continuously Variable Slope Delta (CVSD) modulation and $\mu255$ Law Pulse Coded Modulation (PCM). Within each bandwidth, four of the modes result from using a different serial bit-rate for the CVSD technique and the fifth mode is the $\mu255$ Law PCM technique. All ten modes are listed in Table 1; the CVSD modes are distinguished by the serial bit-rate used.

The function of the Buffer board is to temporarily store digitized audio data; both audio data orginating at its own station (which is to be transmitted onto the bus) and audio data orginating at another station (which has been received from the bus) are stored on this board. The Buffer board handles all audio data in eight-bit bytes. The audio data to be transmitted onto the bus is accumulated one byte at a time at a constant rate. This rate is determined from the selected voice digitization mode. When enough bytes have been accumulated for transmission as a block on the MIL-STD-1553 bus, the 1553 board is so notified. The accumulated audio data is sent upon request to the 1553 board at the bus rate (two bytes every twenty microseconds).

The audio data which is received from the bus is stored in a portion of the memory on the Buffer board dedicated to the station from which the data originated. At the same rate at which a byte of data is accumulated for transmission purposes, the Buffer board sends a byte of data to the Analog board from each portion of its memory where data received from the bus has been stored. In this way the Buffer board reconciles the relatively slow analog data rate with the fast MIL-STD-1553 data rate.

The function of the 1553 board is to transmit data onto the bus and to receive data from the bus utilizing the protocol of MIL-STD-1553. In addition to the audio data, the 1553 board is concerned with control information. The two types of control information present in the MUDIV system are the voice digitization mode and the currently active communication paths. The control information is interleaved with the

FRONT PANEL I/O CONNECTIONS

```
                                    ┌──────────────┐     ┌──────────────┐
                                    │   BUFFER     │     │  INTERFACE   │
                                    │  CONTROLLER  │     │  CONTROLLER  │
                                    └──────────────┘     └──────────────┘

                    ┌──────────────┐   ┌──────────────┐
MICROPHONE INPUTS ─→│     A/D      │──→│ XMIT MEMORY  │────→┌──────────────┐
                    │  CONVERTER   │   │              │     │              │   MIL-STD-1553
                    └──────────────┘   └──────────────┘     │  MIL-STD-1553│ ←→  DATA BUS
                                                            │   INTERFACE  │
   OUTPUT          ┌──────────────┐   ┌──────────────┐     │              │
     TO        ←───│     D/A      │←──│     RCV      │←────│              │
  HEADPHONES       │  CONVERTER   │   │   MEMORY     │     └──────────────┘
                   └──────────────┘   └──────────────┘

     ANALOG BOARD          BUFFER BOARD        MIL-STD-1553 BOARD
```
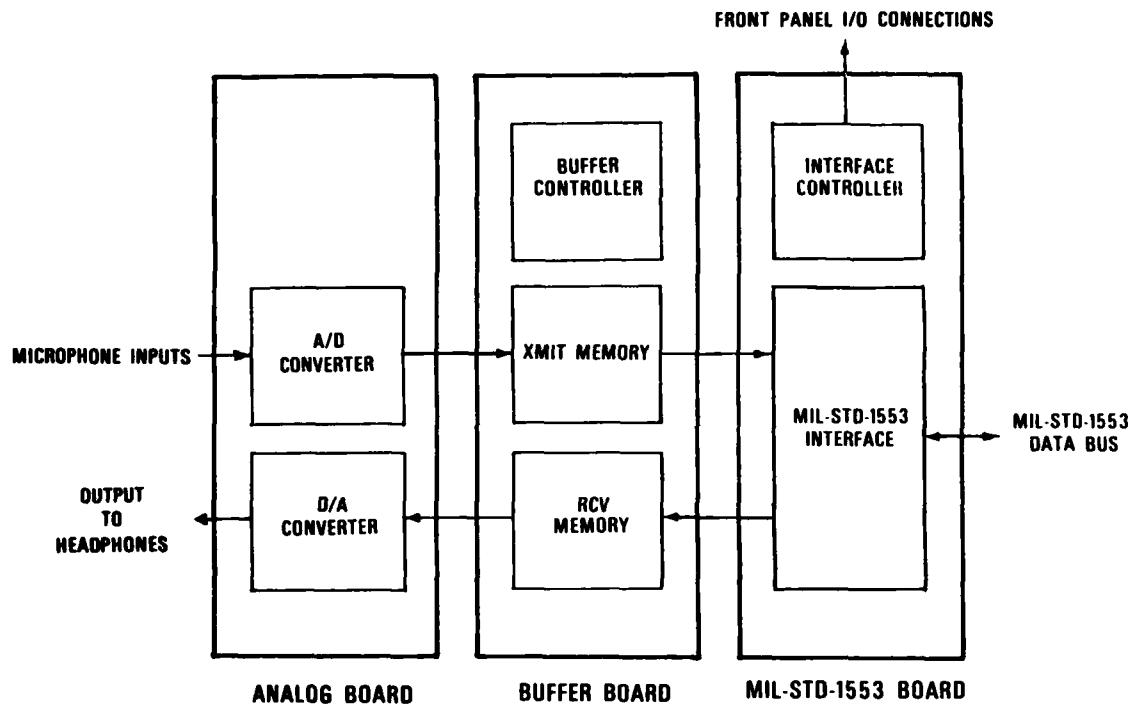
Figure 1.   MUDIV ICS STATION BLOCK DIAGRAM

Table 1.   VOICE CONVERSION MODES

Voice Conversion Techniques	Audio (KHz)	Bandwidth
µ255 Law PCM	3.5	4.5
50 Kbit/sec CVSD	3.5	4.5
32 Kbit/sec CVSD	3.5	4.5
25 Kbit/sec CVSD	3.5	4.5
16 Kbit/sec CVSD	3.5	4.5

audio data on the bus. The 1553 board sorts the control information from the audio data and treats each appropriately.

In order assure that the next generation ICS does not degrade intelligibility, the MUDIV system was used to measure the intelligibility of all ten digitization modes in four different acoustic noise environments. Three of the noise environments were reconstructed from measurements made during flights in U.S. military aircraft; the fourth noise environment was a quiet reference condition. The three aircraft noise environments were: 1) a high performance jet fighter (F-15A), 2) a multi-engine turbo-prop aircraft (EC-130E), and 3) a large helicopter (HH-53C). These intelligibility tests were conducted at the Voice Communication Research and Evaluation System (VOCRES) facility of the Aerospace Medical Research Laboratory (AMRL) at Wright-Patterson Air Force Base (WPAFB), Dayton, Ohio. Each digitization mode and acoustic noise combination was tested using the Modified Rhyme Test.

To insure an accurate simulation of the acoustic environment, the test subjects used headgear representative of that used on actual aircraft. For the F-15A tests, oxygen masks with the Air Force M-101/AIC microphone element and Air Force helmets with H-143 headphone elements were used. For the EC-130E and quiet reference tests, boom microphones with the Navy M-95B/UR microphone element and Navy H-173 headsets were used. For the HH-53C tests, Navy boom microphones and Air Force helmets were used. The preliminary results of these intelligibility tests will be discussed later.

SOLUTIONS OFFERED BY A MUDIV ARCHITECTURE

An ICS based on the MUDIV architecture offers solutions to current ICS problems without imposing new problems on the Fleet. The performance problem of insufficient isolation of an ICS channel from EMI sources can be solved by properly encoding the information before it has been affected by EMI. Relative to an individual ICS channel, all other ICS channels represent noise sources. Electro-Magnetic energy conducted or radiated into an ICS station or the wiring between stations also represents a noise source. In the MUDIV design, the audio signal is converted into digital form as soon as possible after it enters a station. Careful layout of components and shielding of signals are required in order to keep the effects of noise on the audio signal at acceptable levels until the audio signal has been converted into digital data. Once the audio signal is in digital form, layout and shielding rules may be partially relaxed. MIL-STD-1553B requires that terminals be inter-connected with a twisted shielded pair of wires (usually called the bus). MIL-STD-1553B also requires the two bus wires to be driven differentially. For conditions stated in MIL-STD-1553B, the word error rate for a terminal must be no greater than one error in 10,000,000 words received by that terminal. Thus, error correcting codes are not normally required to correct the adverse effects of EMI.

The performance problem of insufficient flexibility in station selection and system configuration can be solved by logically interconnecting ICS stations rather than just physically interconnecting them. MIL-STD-1553B requires that no station transmit or receive any data without first being commanded by the bus controller. Thus, by issuing the proper set of commands, the bus controller can allow any set of stations to communicate with any other set of stations. This degree of flexibility in station selection and system configuration is not practical in the distributed analog type of architecture. The bundle of shielded wires that would be required to physically interconnect the ICS stations in all possible ways would be too large and too heavy for use in aircraft. This degree of flexibility in station selection and system configuration has not been incorporated into a central switch type of architecture for two reasons. First, there has been no requirement to design an ICS that is compatible with multiple aircraft types. Second, the level of integration of electronic components has not been high enough to make an ICS with such flexibility attractive considering the penalties of size, weight, and power consumption.

ICS Availability Problem Solution

ICS equipment availability problems are generally due to the difficulty of repairing the ICS and the logistics problems of supporting the number of different ICS types in the Fleet. The AVCS ICS will be easier to repair because of the design's modularity and its Built-In-Test (BIT) features. Current ICS are difficult to repair due to the lack of BIT. Thus, a skilled technician is required to understand how the circuits function, troubleshoot the equipment, identify and replace faulty components. The distributed analog type of architecture has an additional problem due to the mass of wiring used to interconnect the ICS stations. This wiring is not easily accessible and it may be routed differently than indicated on the aircraft wiring diagram. These difficulties complicate the job of repairing the ICS.

The AVCS ICS design is modular in both mechanical and electronic functional aspects, which are applied at two different levels. These two levels are the Weapon Replaceable Assembly (WRA or "black box") level and the Shop Replaceable Assembly (SRA or "circuit card") level. The BIT features of the AVCS ICS will take advantage of both levels of modularity. The system portion of the BIT will be directed toward identifying a faulty WRA. The local portion of the BIT (within a particular WRA) will be directed toward identifying the set of potentially faulty SRA's that could cause the observed error in the WRA. The BIT will operate on electronic functional groups of circuits. The high degree of correlation between the electronic functional circuits and the mechanical modules eases the job of repairing the AVCS ICS. The new ICS can be repaired by activating the BIT and replacing the indicated SRA's. Since twisted shielded pairs of wire interconnect ICS stations, maintenance personnel will require minimum systems knowledge to troubleshoot the part not covered by BIT, i.e., the interconnecting wiring must be intact. Thus, the AVCS ICS will be easier to repair, and can be repaired by less skilled personnel.

Since most of the AVCS ICS assemblies will be interchangeable from station-to-station within in aircraft as well as from aircraft-to-aircraft, the logistics problems of field support will be reduced. Thus, the spare parts needed to repair the AVCS ICS will be more readily available. Parts obsolescence contributes to the problem of supplying spare parts for current ICS. The modularity aspects of the AVCS ICS design combat the obsolescence problem. As parts used for a certain function become obsolete, new parts, able to perform the same function, can be substituted into the system without major rework or redesign. For example, the voice digitization components chosen for the AVCS ICS may be superseded in the near future. The AVCS ICS design anticipates this by placing the "Analog board" function (of the MUDIV system) on a few SRA's. Thus, by changing a few SRA's the voice digitization technique used by the ICS can be changed. Because the mechanical modularity is highly correlated with the functional modularity, a change such as this can be easily accomplished in the field.

Potential Problems Averted

Two potential problems in the development of a new ICS have been averted by the AVCS ICS. These problems are the intelligibility of the system and the ability of all potential users to incorporate this system into their aircraft. To insure the suitability of the AVCS ICS in operational situations, the intelligibility of the voice digitization modes under consideration have been tested under simulated operating conditions. In order to develop an ICS usable by all three branches of the U.S. military, the AVCS ICS development efforts have been coordinated among the three services. The intelligibility test results and Tri-Service coordination efforts are described below.

Intelligibility Test Results

The preliminary intelligibility test results are presented in Figures 2 and 3. Both figures show the intelligibility at five digital bit-rates under four ambient noise conditions. Figure 2 presents results for the five modes with a 3500 Hz analog audio bandwidth. Figure 3 presents the results for the five modes with a 4500 Hz analog audio bandwidth. In all ten modes the level of ambient noise markedly effects the intelligibility. Generally the intelligibility in the 4500 Hz analog audio bandwidth modes is better than the intelligibility in the 3500 Hz analog audio bandwidth modes. In the highest ambient noise case tested, the 50,000 bit/sec CVSD technique yielded better intelligibility than the higher bit-rate μ255 Law PCM technique. Also in the highest ambient noise case, the 16,000 bit/sec CVSD technique yielded unacceptable intelligibility. Based on these results, the 32,000 bit/sec CVSD technique with an analog audio bandwidth of 4500 Hz is the lowest bit-rate mode with the intelligibility required for military ICS.
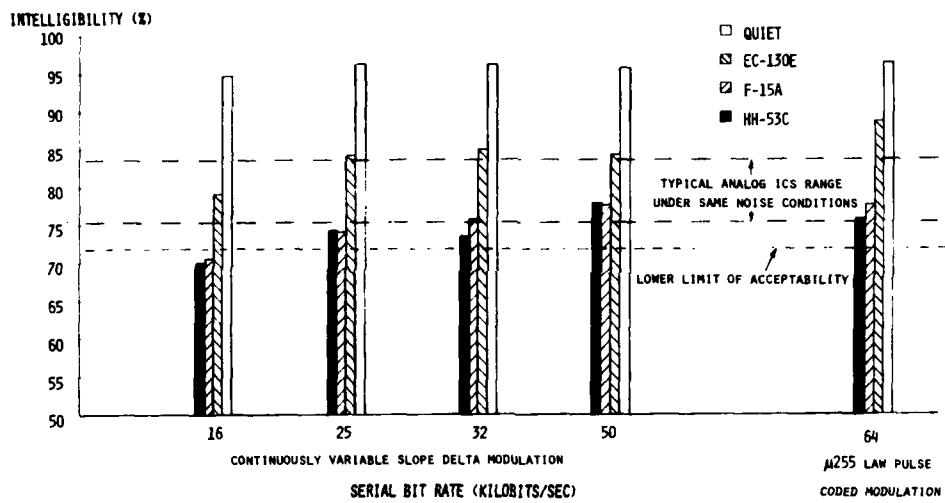

Figure 2. Intelligibility at Selected Bit Rates (3500 Hz Analog Audio Bandwidth)
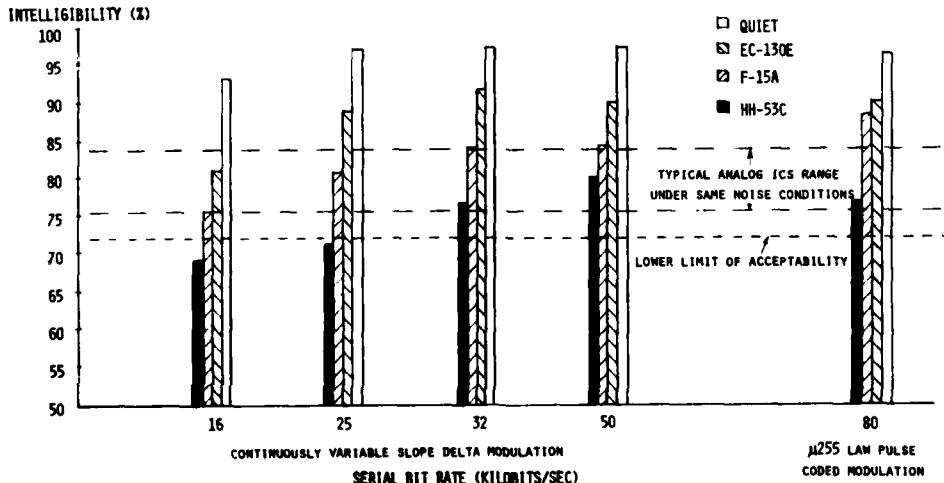

Figure 3. Intelligibility at Selected Bit Rates (4500 Hz Analog Audio Bandwidth)

Tri-Service Coordination Efforts

In order to be used on a number of different aircraft, the AVCS ICS must meet the needs of all applicable aircraft. The current ICS development is coordinating activities among the U.S. Air Force, Army, and Navy in an effort to develop an ICS useful to all three services. An example of this coordination is the incorporation of the 4500 Hz analog audio bandwidth modes into the MUDIV system at the Army's suggestion. A Memorandum of Agreement (MOA) on general avionics standardization has been signed by the Assistant Secretaries of the Air Force, Army, and Navy. An MOA on Tri-Service ICS development is planned for 1981. Following this specific MOA on ICS development, a draft specification for a Tri-Service ICS will be prepared and submitted to all three services and industry for review.

CONCLUSION

Existing Intercommunication Systems have problems that should be resolved in the next generation ICS. The major performance problems are a lack of isolation between stations and a lack of flexibility in station selection and system configuration. The major availability problems are related to maintenance and logistics due to the lack of equipment standardization.

The major problems of the ICS can be solved by the proper use of state-of-the-art technology. The performance problems can be solved by distributing intelligent stations throughout the aircraft, and passing information between the stations in digital form via a fault tolerant bus network. The availability problems can be solved by ICS standardization and by a mechanical and electronic functional modular design that is self-testing and used on multiple aircraft. This modular design will allow the incorporation of future voice conversion techniques without major rework or redesign.

RECOMMENDATIONS

Based on the AVCS ICS development effort, the following recommendations are offered as follow-on tasks:

1. Measure the intelligibility of low bit rate (less than 16,000 bit/sec) voice digitizing techniques in aircraft acoustic noise environments

2. Integrate voice synthesis capability into the MUDIV ICS for use in testing the effectiveness of aural warnings

3. Integrate the presently separate ICS and radio controls into a common set of communication controls.

REFERENCES

1. R. F. Bolt and B. D. Sanderson, "A Multiplexed Digital Voice Intercommunications System for Military Applications", SAE Technical Paper #801144, October 1980.

2. G. R. McClure, and W. R. Sobko "General Purpose Audio-Multiplexed Intercom System" Proceedings of 2nd AFSC Multiplex Data Bus Conference, October 1978.

3. C. Teacher and H. Watkins "ANDVT Microphone and Audio System Study" Final Report to Commander Naval Electronic System Command under Contract No. N00039-77-C-0365, August 1978.

4. M. Minnich, "Test Report: BASIC Laboratory Evaluation of Data Bus Voice Multiplexed System" prepared for BASIC Laboratory, Naval Air Development Center, Warminister, PA 18974, February 1930.

5. MIL-STD-1553B "Aircraft Internal Time Division Command/Response Multiplex Data Bus", 30 April 1975.

6. W. A. Crossgrove, D. H. Ellis, et al "Multiplex Applications Handbook", The Boeing Company, Seattle, WA, and SCI Systems, Inc., Huntsville, AL, May 1979.

SIDETONE-LEVEL CONSIDERATIONS IN AIRCRAFT COMMUNICATION SYSTEMS

S.E. Forshaw
Behavioral Sciences Division
Defence and Civil Institute of Environmental Medicine
1133 Sheppard Avenue West
Downsview, Ontario, M3M 3B9
CANADA

SUMMARY

It is well known that the loudness of an individual's speech is governed by two possibly related feedback phenomena: the "sidetone-amplification effect" - the tendency for a speaker to decrease his vocal effort when he hears his voice at an amplified level, and the "Lombard effect" - the tendency for a speaker to increase his vocal intensity in the presence of noise. It has been noted previously that the speech intelligibility characteristics of some military aircraft communications systems require improvement. The limited noise-cancelling capabilities of current microphones have been identified, for example, as the major source of system-amplified noise reaching listeners ears. It may be possible to reduce the level of this source of noise, and hence increase a system's signal-to-noise ratio and speech intelligibility, by optimizing the level of a speaker's transmit sidetones to elicit an increase in his vocal output.

INTRODUCTION

Aircrew flying the Canadian Forces (CF) CH-147 Chinook helicopter have reported difficulty in hearing and understanding the transmit voice-signal (sidetone) of the AN/ARC-114 VHF/FM, ARC-115 VHF/AM and ARC-116 UHF/AM receiver/transmitters over the background noise of the aircraft and its systems. Measurement of the audio signal levels in the aircraft's C-6533/ARC Intercommunication Control System (ICS) led to the conclusion that the problem was not due to defects in the equipment per se, but to :he relative settings of the transmit-sidetone and received-audio signal levels (1).

In each of these receiver/transmitters, the two signals share a common volume control and output amplifier. When in transmit mode, a switch within the transmitter removes the received-audio signal from the amplifier lines and replaces it with a transmit-sidetone signal. However, a loss of transmitter power, low output power, or a high voltage-standing-wave ratio disables the transmit-sidetone signal and is intended to indicate to an operator that the transmitter is not functioning. The difference in level between the received-audio and transmit-sidetone signals is adjusted prior to installation of the equipment in the aircraft. The manufacturer-installed resistor network allows the sidetone to be set from 3 to 10 dB below the received-audio level.

Transmit-sidetone levels were being preset typically 7 to 10 dB below received-audio signals, depending on the frequency selected. This difference, about two-to-one in loudness, was acceptable in hangar tests when only electrical systems were turned on in the aircraft and the background noise was relatively low. Before takeoff, aircrew would adjust their intercom volume controls so that the received-audio signals could be heard at a comfortable level. The resulting transmit-sidetones were barely audible in the ambient noise of the aircraft in flight.

It was therefore recommended that the transmit-sidetone levels be set no more than 3 dB below the received-audio signals (1). To facilitate this adjustment, it was necessary to replace the manufacturer's fixed-resistor network in the transmit-sidetone level control with a variable potentiometer. In subsequent flight trials, aircrews judged the "3-dB down" transmit-sidetone levels to be a substantial improvement (2).

PSYCHOACOUSTIC CONSIDERATIONS

The sidetone-level problem encountered in the CH-147 helicopter prompted the author to review the psychoacoustic factors influencing the effectiveness of voice communications in intense-noise environments. It is well known that the loudness of an individual's speech, and hence the signal-to-noise ratio at his microphone, is governed by two possibly related feedback phenomena: the "sidetone-amplification-effect" - the tendency for a speaker to lecrease his vocal effort when he hears his voice at an amplified level (3), and the "Lombard effect" - the tendency for a speaker to increase his vocal intensity in the presence of noise (4).

In early sidetone-amplification studies, the intensity of the noise at the speakers' ears was not a variable. In the experiments of Lane, Tranel and Sisson (5), for example, listening conditions were intended to simulate those of an aircraft; hence speakers listened to their sidetones in a constant 75-dB sound pressure level (SPL) background noise.

More recently, however, Siegel and Pick (6) demonstrated that the level of the noise at a speaker's ears affects the sidetone-amplification effect. Up to 80 dB SPL, at least, the more intense the "speech-spectrum" masking noise, the greater the effect (see Figure 1). In the 0-dB noise condition, speakers' mean vocal output increased by 2.2 dB as sidetone amplification was reduced from 20 to 0 dB, whereas in the 80-dB noise condition, the output increased by 6.8 dB.

The Lombard effect is also evident in Figure 1. In the 0-dB sidetone-amplification condition, the speakers' vocal output increased by about 6 dB as the "speech-spectrum" masking noise in the earphones increased from 60 to 80 dB SPL.

Garber et al (7) demonstrated that the spectrum of the masking noise is a significant factor affecting both the Lombard and the sidetone-amplification effects. Three bands of white noise (20-20000, 1800-2500, and 4000-6000 Hz), each adjusted to an overall SPL of 100 dB, were used as a background noise in which the vocal output of speakers was monitored relative to their vocal output in quiet. The mean increases in speech levels in the three types of noise were 15,11, and 7 dB respectively (see Figure 2).

Subjects also spoke in the presence of the three bands of noise for two sidetone conditions: 0- and 20-dB amplification. The mean changes in the speakers' vocal output (the difference in vocal inten- sities between the 0- and 20-dB amplification conditions) were 6.7, 5.6, and 4.9 dB respectively (see Figure 2). However, it was noted that only the difference in vocal-output changes between the 20-20000 and the 4000-6000 Hz bandwidth conditions were statistically significant (7).

The three bands of noise were also used to mask phonetically-balanced word lists at a signal-to- noise ratio of -13 dB. The mean speech-reception intelligibility scores achieved by a group of normal- hearing listeners were 0, 29, and 80 per cent correct for masking-noise bandwidths of 20-20000, 1800-2500, and 4000-6000 Hz respectively (see Figure 2). When the same bands of noise were adjusted for equal loudness, the mean-intelligibility scores followed the same trend: 10, 22, and 72 per cent correct respectively.

Not surprisingly, the Lombard effect is related to the intelligibility of a talker's speech. Noises that degrade intelligibility the most elicit the largest effect; noises that degrade the least induce the smallest effect. The relationship between the sidetone-amplification effect and masking effectiveness does not appear to be as pronounced. In quiet, a normal-hearing speaker hears his own voice by means of air- and bone-conducted signals, the air-conducted component being about 6 dB louder (8). If the air-conducted signal is amplified, thus producing an extraordinarily loud sidetone, the speaker reduces his vocal effort accordingly. However, the reduction is minimal; at most, a speaker adjusts his voice level by about 1 dB for a change of 2 dB in amplified sidetone level (3).

In noise, a normal-hearing speaker with amplified sidetone receives masked bone-conducted and air- conducted signals. The amplified air-conducted signal becomes the speaker's best cue for monitoring his voice, as long as he expends sufficient vocal effort to maintain the sidetone at some minimum level. As the masking effectiveness of the ambient noise increases, the need for vocal effort is reduced as side- tone amplification is increased.

DISCUSSION

In vew of the above results, a point to be considered is whether some optimization of aircraft intercom transmit-sidetone levels might yield improvements in speech intelligibility. Two questions must be addressed in this regard: Does the increase in signal-to-noise ratio realized at a speaker's micro- phone because of extra vocal effort translate into increased speech intelligibility? In operational fly- ing, would an intercom system with significantly reduced sidetone amplification elicit increased vocal effort from, and be acceptable to the user?

Pickett reports that up to about 15 to 20 dB above normal voice level, speech intelligibility drops by not more than 5 per cent. Above this level, however, there is a considerable reduction in intelligibility, particularly for consonant sounds and for words spoken without phrase or sentence con- text (9,10).

To increase the vocal output of speakers in the extremely intense noise environment of aircraft launch and recovery operations, sidetone-amplification was not provided in flight-deck radio helmets aboard the aircraft carrier USS LEXINGTON (11). In this instance, the lack of amplified sidetone did not appear to bother users of the system. It was not clear, however, whether the absence of the sidetone had the intended effect of increasing vocal output.

Klumpp and Webster (12) suggested that a major factor in determining a speaker's output level is his knowledge of the vocal effort required from previously successful communications. Similarly, Lane and Tranel (4) have argued that a speaker's perception of difficulties in his communication link and the importance he attaches to the successful reception of the message preempt control of his vocal output.

It is clear that pilots' speech transmissions were not loud enough to be heard or understood by other members of the crew in CF CH-147 helicopters when the transmit-sidetone levels were set 7 to 10 dB below received-audio signals. Presumably, pilots were able to hear their own amplified sidetones to be able to monitor transmitter operations. However, a vocal effort sufficient to provide a just-audible amplified sidetone to a pilot in the cockpit of the CH-147 is not likely to be loud enough for listeners in noisier areas of the aircraft.

Undoubtedly, the automatic gain control (AGC) in the CH-147 ICS microphone amplifier precludes any significant sidetone-amplification effect. Of course, AGC is essential to maintain a required level of carrier modulation. However, if the speaker's sidetone signal is also controlled by the AGC circuit, as it is in the CH-147 receiver/transmitters, the AGC will tend to maintain a constant level of amplified sidetone in the speaker's headset, irrespective of vocal output. As suggested previously, it is natural in such a situation for a speaker to reduce his vocal output.

It has been noted in earlier AGARD papers that the speech intelligibility characteristics of at least some military-aircraft communication systems must be improved to enhance operational effectiveness.

and increase flight safety. The limited far-field (noise-cancelling) characteristics of the current M-87 microphone have been identified, for example, as the major source of ICS-amplified noise reaching speakers' and listeners' ears in the Chinook helicopter (13,14).

Yet it has been noted that replacement of the M-87 with a microphone possessing "ideal" far-field response characteristics, together with an improved transmission-whine filter in the ICS circuitry, would only reduce the amplified noise reaching listeners' ears to a SPL of about 92 dBA (15), still 5 or 6 dB above the average noise level penetrating a pilot's SPH-4 or CF 411 flight helmet (2,14).

Perhaps it is possible to reduce this ICS processed noise even further, and at the same time increase the system intelligibility, by optimizing the speaker's transmit-sidetone level to elicit an increase in vocal level and signal-to-noise ratio at his microphone.

RECOMMENDATIONS

It is recommended that in the design of future aircraft-communication systems, consideration be given to optimizing the level of the transmit-sidetone signal that is fed back to speakers' headsets. It would require that this signal be free from the action of any AGC circuitry if advantage is to be obtained in signal-to-noise ratio at the speaker's microphone from the Lombard and sidetone-amplification effects.

REFERENCES

1. Moore, R.M. Canadian Forces Aerospace Engineering Test Establishment, Cold Lake, Alberta. Investigation of Audio Communications Problems CH-147 Aircraft. 1976. AETE Report 76/26.

2. Forshaw, S.E. Defence and Civil Institute of Environmental Medicine, Downsview, Ontario. CH-147 Modified Standard Light Avionics Equipment (SLAE) Trials. 1980. DCIEM Technical Memorandum 3615A10.

3. Lane, H.E., Catania, A.C. and Stevens, S.S. Voice Level: Autophonic Scale, Perceived Loudness, and Effects of Sidetone. Jour. Acoust. Soc. Am. 33: 160-167, 1961.

4. Lane, H. and Tranel, B. The Lombard Sign and the Role of Hearing in Speech. Jour. Sp. Hearing Res. 14: 677-709, 1971.

5. Lane, H., Tranel, B. and Sisson, C. Regulation of Voice Communication by Sensory Dynamics. Jour. Acoust. Soc. Am. 47: 618-624, 1970.

6. Siegel, G.M. and Pick, H.L. Auditory Feedback in the Regulation of Voice. Jour. Acoust. Soc. Am. 56: 1618-1624, 1974.

7. Garber, S.F., Siegel, G.M., Pick, H.L. and Alcorn, S.R. The Influence of Selected Masking Noises on Lombard and Sidetone Amplification Effects. Jour. Sp. Hearing Res. 19: 523-535, 1976.

8. von Bekesy, G. Experiments in Hearing. New York, McGraw-Hill Book Company Ltd., pp. 183-187, 1960.

9. Picket, J.M. Effects of Vocal Force on the Intelligibility of Speech Sounds. Jour. Acoust. Soc. Am. 28: 902-905, 1956.

10. Pickett, J.M. Limits of Direct Speech Communication in Noise. Jour. Acoust. Soc. Am. 30: 278-281, 1958.

11. Webster, J.C. and Klumpp, R.G. U.S. Navy Electronics Laboratory, San Diego, California. Final Report: USNEL Flight Deck Communications System. Part 2. Noise and Acoustic Aspects. NEL Report 923, 1960.

12. Klumpp, R.G. and Webster, J.C. U.S. Navy Electronics Laboratory, San Diego, California. Final Report: USNEL Flight Deck Communications System. Part 5. Message Analysis and Personnel Response. NEL Report 926, 1960.

13. Mayer, M.S. and Lindberg, A.W. A Survey of Communications in the High Noise Environment of Army Aircraft. AGARD Conference Proceedings No. 255, 1978.

14. Camp, R.T. and Mozo, B.T. An Investigation of Aircraft Voice Communication Systems as Sources of Insidious Long-Term Acoustic Hazards. AGARD Conference Proceedings No. 255, 1978.

15. Giordano, T.A. and Keane, G.C. Communications/ADP Laboratory, Army Electronic Command, Fort Monmouth, N.J. The Effect of Helicopter Noise on Communication and Hearing. Research and Development Technical Report ECOM-4140, 1973.
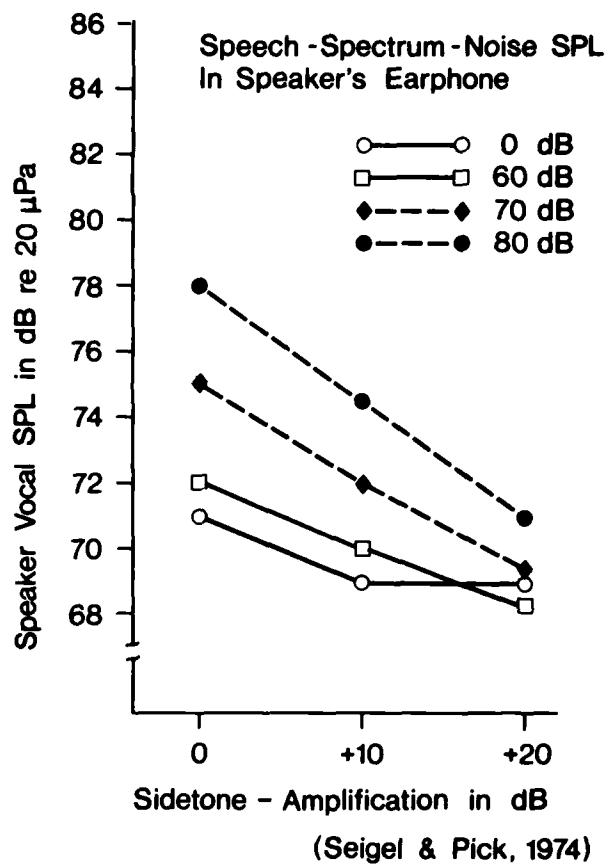
FIGURE 1. Effects of masking noise and sidetone amplification
upon speaker vocal output.

(Garber et al, 1976)
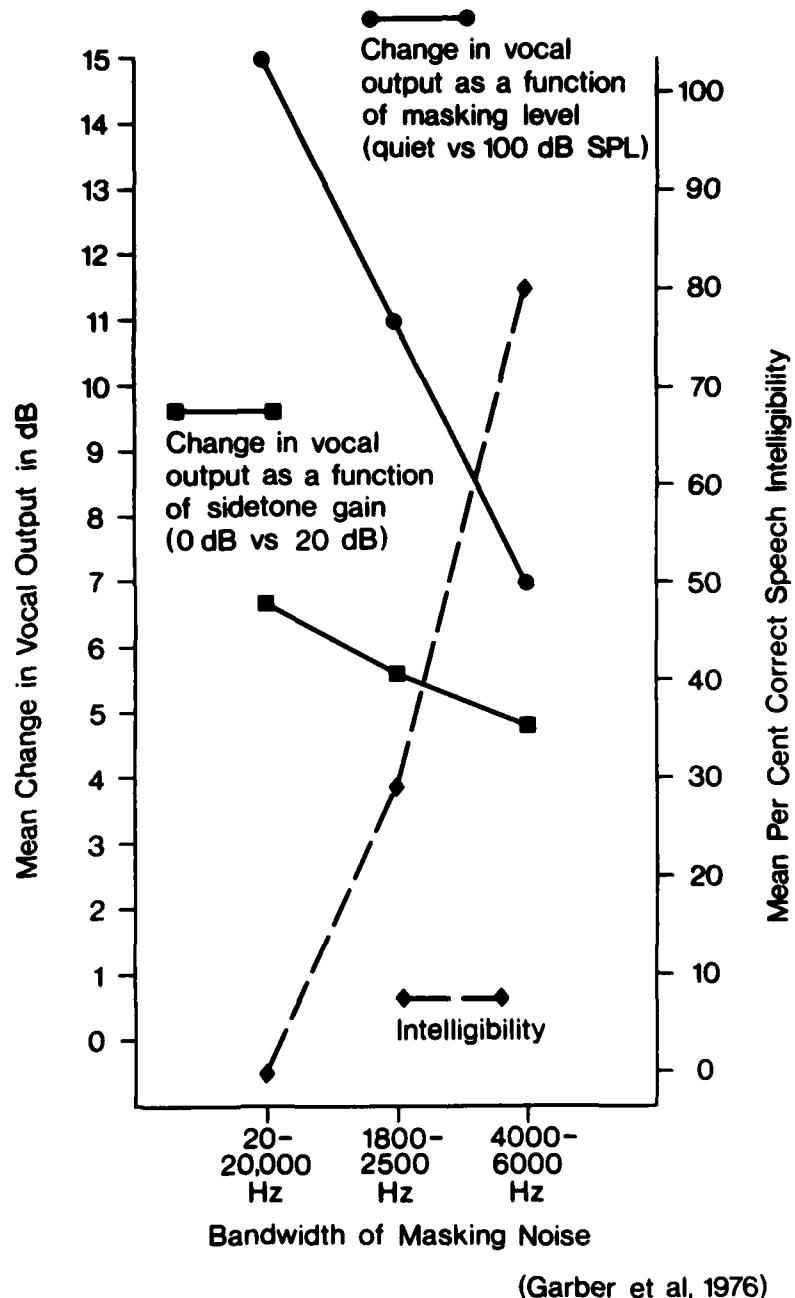
FIGURE 2. Effects of bandwidth of masking noise upon speech
intelligibility, speaker vocal effort, and sidetone-
amplification effect.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
	AGARD-CP-311	ISBN 92-835-1388-6	UNCLASSIFIED

5. Originator
Advisory Group for Aerospace Research and Development
North Atlantic Treaty Organization
7 rue Ancelle, 92200 Neuilly sur Seine, France

6. Title

AURAL COMMUNICATION IN AVIATION

7. Presented at

the Aerospace Medical Panel Specialists' Meeting held in Soesterberg, Netherlands 30 March—2 April 1981.

8. Author(s)/Editor(s)	9. Date
Dr K.E.Money	June 1981

10. Author's/Editor's Address Defence and Civil Institute of Environmental Medicine, 1133 Sheppard Avenue West, P.O. Box 2000, Downsview, Ontario M3M 3B9, Canada	11. Pages 196

12. Distribution Statement   This document is distributed in accordance with AGARD policies and regulations, which are outlined on the Outside Back Covers of all AGARD publications.

13. Keywords/Descriptors

Voice communication	Military aircraft
Auditory perception	
Intelligibility	

14. Abstract

Despite the dependance of military operations in air, on land or sea on reliable voice communication and the effective use of audio warnings, many of the systems currently in use have serious shortcomings and do not reflect the considerable research effort that has been expended.

In modern military aircraft, it is essential that aircrew should be able to perceive and respond to audio information with minimum effort and highest reliability. However, the low quality of most airborne voice communications systems imposes such a high additional workload that messages are liable to misinterpretation or to being missed altogether.

Hearing standards and conservation are also discussed.

AGARD-CP-311

Voice communication
Auditory perception
Intelligibility
Military aircraft

AGARD Conference Proceedings No.311
Advisory Group for Aerospace Research and Development, NATO
AURAL COMMUNICATION IN AVIATION
Edited by Dr K.E.Money
Published June 1981
196 pages

Despite the dependance of military operations in air, on land or sea on reliable voice communication and the effective use of audio warnings, many of the systems currently in use have serious shortcomings and do not reflect the considerable research effort that has been expended.

In modern military aircraft, it is essential that aircrew should be able to perceive and respond to audio

P.T.O.

AGARD-CP-311

Voice communication
Auditory perception
Intelligibility
Military aircraft

AGARD Conference Proceedings No.311
Advisory Group for Aerospace Research and Development, NATO
AURAL COMMUNICATION IN AVIATION
Edited by Dr K.E.Money
Published June 1981
196 pages

Despite the dependance of military operations in air, on land or sea on reliable voice communication and the effective use of audio warnings, many of the systems currently in use have serious shortcomings and do not reflect the considerable research effort that has been expended.

In modern military aircraft, it is essential that aircrew should be able to perceive and respond to audio

P.T.O.

AGARD-CP-311

Voice communication
Auditory perception
Intelligibility
Military aircraft

AGARD Conference Proceedings No.311
Advisory Group for Aerospace Research and Development, NATO
AURAL COMMUNICATION IN AVIATION
Edited by Dr K.E.Money
Published June 1981
196 pages

Despite the dependance of military operations in air, on land or sea on reliable voice communication and the effective use of audio warnings, many of the systems currently in use have serious shortcomings and do not reflect the considerable research effort that has been expended.

In modern military aircraft, it is essential that aircrew should be able to perceive and respond to audio

P.T.O.

AGARD-CP-311

Voice communication
Auditory perception
Intelligibility
Military aircraft

AGARD Conference Proceedings No.311
Advisory Group for Aerospace Research and Development, NATO
AURAL COMMUNICATION IN AVIATION
Edited by Dr K.E.Money
Published June 1981
196 pages

Despite the dependance of military operations in air, on land or sea on reliable voice communication and the effective use of audio warnings, many of the systems currently in use have serious shortcomings and do not reflect the considerable research effort that has been expended.

In modern military aircraft, it is essential that aircrew should be able to perceive and respond to audio

P.T.O.

information with minimum effort and highest reliability. However, the low quality of most airborne voice communications systems imposes such a high additional workload that messages are liable to misinterpretation or to being missed altogether.

Hearing standards and conservation are also discussed.

Papers presented at the Specialists' Meeting of the Aerospace Medical Panel held at Soesterberg, Netherlands 30 March–2 April 1981.

---

information with minimum effort and highest reliability. However, the low quality of most airborne voice communications systems imposes such a high additional workload that messages are liable to misinterpretation or to being missed altogether.

Hearing standards and conservation are also discussed.

Papers presented at the Specialists' Meeting of the Aerospace Medical Panel held at Soesterberg, Netherlands 30 March–2 April 1981.

AGARD

NATO ⊕ OTAN

7 RUE ANCELLE · 92200 NEUILLY-SUR-SEINE
FRANCE

Telephone 745.08.10 · Telex 610176

DISTRIBUTION OF UNCLASSIFIED
AGARD PUBLICATIONS

AGARD does NOT hold stocks of AGARD publications at the above address for general distribution. Initial distribution of AGARD publications is made to AGARD Member Nations through the following National Distribution Centres. Further copies are sometimes available from these Centres, but if not may be purchased in Microfiche or Photocopy form from the Purchase Agencies listed below.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordonnateur AGARD – VSL
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA
Defence Science Information Services
Department of National Defence
Ottawa, Ontario K1A OK2

DENMARK
Danish Defence Research Board
Østerbrogades Kaserne
Copenhagen Ø

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92320 Châtillon sous Bagneux

GERMANY
Fachinformationszentrum Energie,
Physik, Mathematik GmbH
Kernforschungszentrum
D-7514 Eggenstein-Leopoldshafen 2

GREECE
Hellenic Air Force General Staff
Research and Development Directorate
Holargos, Athens

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
3, Piazzale Adenauer
Roma/EUR

LUXEMBOURG
See Belgium

NETHERLANDS
Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
P.O. Box 126
2600 A.C. Delft

NORWAY
Norwegian Defence Research Establishment
Main Library
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Direcção do Serviço de Material
da Forca Aerea
Rua da Escola Politécnica 42
Lisboa
Attn: AGARD National Delegate

TURKEY
Department of Research and Development (ARGE)
Ministry of National Defence, Ankara

UNITED KINGDOM
Defence Research Information Centre
Station Square House
St. Mary Cray
Orpington, Kent BR5 3RE

UNITED STATES
National Aeronautics and Space Administration (NASA)
Langley Field, Virginia 23365
Attn: Report Distribution and Storage Unit

THE UNITED STATES NATIONAL DISTRIBUTION CENTRE (NASA) DOES NOT HOLD
STOCKS OF AGARD PUBLICATIONS, AND APPLICATIONS FOR COPIES SHOULD BE MADE
DIRECT TO THE NATIONAL TECHNICAL INFORMATION SERVICE (NTIS) AT THE ADDRESS BELOW.

PURCHASE AGENCIES

Microfiche or Photocopy	*Microfiche*	*Microfiche*
National Technical Information Service (NTIS) 5285 Port Royal Road Springfield Virginia 22161, USA	Space Documentation Service European Space Agency 10, rue Mário Nikis 75015 Paris, France	Technology Reports Centre (DTI) Station Square House St. Mary Cray Orpington, Kent BR5 3RF England

Requests for microfiche or photocopies of AGARD documents should include the AGARD serial number, title, author or editor, and publication date. Requests to NTIS should include the NASA accession report number. Full bibliographical references and abstracts of AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
published by NASA Scientific and Technical
Information Facility
Post Office Box 8757
Baltimore/Washington International Airport
Maryland 21240, USA

Government Reports Announcements (GRA)
published by the National Technical
Information Services, Springfield
Virginia 22161, USA

DATE

ILMED

-8